

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2004, 12:07:52 ; Search time 188 Seconds  
(without alignments)  
11503.424 Million cell updates/sec

Title: US-10-021-657-7  
Perfect score: 3897  
Sequence: 1 gaattccaaaggaggccctt.....gctgagtaaacgaagaattc 3897

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119.6	3.1	1626	3	US-09-158-767-14
2	119.6	3.1	2181	3	US-09-158-767-10
3	65.4	1.7	77536	4	US-09-410-551B-1
4	62.8	1.6	1078	4	US-09-452-239-41
5	59.8	1.5	1926	4	US-09-249-585A-4
6	59.8	1.5	1931	2	US-09-130-114-2
7	58.2	1.5	49377	1	US-08-784-233A-1
8	58	1.5	390	3	US-09-197-649-7
9	57.4	1.5	819	4	US-09-266-965-98
10	57.4	1.5	1018	4	US-09-452-239-35
11	57.4	1.5	3765	4	US-09-266-965-9
12	57.4	1.5	53500	4	US-09-266-965-76
13	57.2	1.5	3794	4	US-09-192-434-1
14	56.4	1.4	3382	2	US-08-682-847-1
15	56	1.4	900	5	PCT-US95-04801-3
16	55.2	1.4	2799	4	US-09-232-279-1
17	55.2	1.4	23673	4	US-09-773-816-1
18	55	1.4	2721	6	5215881-2
19	55	1.4	8438	1	US-07-945-283-1
20	55	1.4	44377	2	US-08-804-227C-7
21	55	1.4	44377	2	US-08-804-198-1
22	54.6	1.4	1110	4	US-09-342-143-1
23	54.6	1.4	1110	4	US-09-924-439-1
24	54.4	1.4	1361	4	US-09-614-912-37
25	54.4	1.4	77536	4	US-09-410-551B-1
26	54	1.4	23673	4	US-09-773-816-1
27	53.6	1.4	4451	4	US-09-548-938A-1

## ALIGNMENTS

### RESULT 1

US-09-158-767-14  
; Sequence 14, Application US/09158767A  
; Patent No. 6180363

GENERAL INFORMATION:

APPLICANT: Batard, Yannick

APPLICANT: Durst, Francis

APPLICANT: Schalk, Michel

APPLICANT: Werck-Reichhart, Daniele

TITLE OF INVENTION: RECORDING OF DNA SEQUENCES PERMITTING

TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST

FILE REFERENCE: A32000

CURRENT APPLICATION NUMBER: US/09/158,767A

CURRENT FILING DATE: 1998-09-23

EARLIER APPLICATION NUMBER: FR 97-12094

EARLIER FILING DATE: 1997-09-24

NUMBER OF SEQ ID NOS: 20

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 14

LENGTH: 1626

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Altered sequences

US-09-158-767-14

Query Match 3.1%; Score 119.6; DB 3; Length 1626;

Best Local Similarity 59.8%; Pred. No. 2.9e-20;

Matches 219; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

QY 2691 CGCGCAGGACCCCAAGGGATCTCGGAGGACGCTGCTCGCGAGCGGACGAGGTGAG 2750

Db 1149 CCCCGAGGATCTCAAGCAGCTGCTCGCGAGGACTACCTCCCGACGACCTTCGTGCC 1208

QY 2751 GCGCGCGGGATGGTACGCTACGTCGCTACTCGATGGGCGGATGGAGTACAACTGGGG 2810

Db 1209 GCGCGGGTGTGGTCACTTCCATATCTACTCGCGGGCGCATGAAGGGGTGTGGGG 1268

QY 2811 CCCCGACGCGGGAGCTTCCGCGCGGAGCGGTGATCAACAGGAGTGGCGC---GTTCCG 2867

Db 1269 GGAGGACTCCCTGAGTTCGCGCGGAGGATGGCTGTGCGCCGACGCGACCAAGTTCGA 1328

QY 2868 CACCGCTCGCGTTCAATTACCGCGTTCACGCGGGCGGCGGAGGATCTGCTCTGGGCAA 2927

Db 1329 GCAGCAGCACTCGTCAAGTTCGTGGCGTTCACGCGGGCGGAGGTTGTCTCTGGGCAA 1388

QY 2928 GCACTCGGCGTACCTGCAGATGAAGTGGCGTGGCCATCTCTTCCGCTTCTACAGCTT 2987

Db 1389 GGACCTAGCTTACCTGCAGATGAAGAACATCGCGGGAGCGTGTCTCTCGGCACGCGCT 1448

Sequence 3, Appli

Sequence 3, Appli

Sequence 3, Appli

Sequence 3, Appli

Sequence 1, Appli

Sequence 1, Appli

Sequence 1, Appli

Sequence 1, Appli

Sequence 11512, A

Sequence 3, Appli

Sequence 11572, A

Sequence 11871, A

Sequence 2, Appli

Sequence 7285, Ap

Sequence 7174, Ap

Sequence 7330, Ap

Sequence 14835, A

QY 2988 CCGGCTGCTGAGGGGACCCGGGTGACATCCGATGATGACCATCTCTCCATGGCGCA 3047  
Db 1449 GACGTTGGCGCGGGCCACCCGGTGGAGAGAGATGTCGCTCAGCTCTTCATGAAGG 1508  
QY 3048 CGGCT 3053  
Db 1509 CGGCT 1514

## RESULT 2

US-09-158-767-10  
; Sequence 10, Application US/09158767A  
; Patent No. 6180363  
; GENERAL INFORMATION:  
; APPLICANT: Bataud, Yannick  
; APPLICANT: Durst, Francis  
; APPLICANT: Schalk, Michel  
; APPLICANT: Werck-Reichhart, Daniele  
; TITLE OF INVENTION: RECORDING OF DNA SEQUENCES PERMITTING  
; TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST  
; FILE REFERENCE: A32000  
; CURRENT APPLICATION NUMBER: US/09/158,767A  
; PRIOR FILING DATE: 1998-09-23  
; EARLIER APPLICATION NUMBER: FR 97-12094  
; PRIOR FILING DATE: 1997-09-24  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Fast-SEQ for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 2181  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
US-09-158-767-10

Query Match 3.1%; Score 119.6; DB 3; Length 2181;  
Best Local Similarity 59.8%; Pred. No. 3.4e-20;  
Matches 219; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

QY 2691 CGCGCAGGACCCCAAGGGATCCTGGAGGACGACGTCTCCCGGAGCGGAGCGAGGTGAG 2750  
Db 1260 CCGGAGGACTCAAGACAGTCTCGGAGACGACTACCTCCCGAGCGGACCTTCGTGCC 1319  
QY 2751 GCGCGCGGGATGTTGACGATGTCGCTACATCGATGGGCGGATGGAGTCAACTGGGG 2810  
Db 1320 GCGCGGGTGTCTGCTCAGCTACCTATATATCTCGGCGGGCGCATGAAGGGGTGGGG 1379  
QY 2811 CCGGAGCGCGGAGCTTCGCGCGGAGCGGTGGATCAAGAGATGGGC-----GTTCCG 2867  
Db 1380 GGAGGACTGCTGAGTTCGCGCGGAGCGATGGCTGTCCGCGGAGCGGACCAAGTTGCA 1439  
QY 2868 CAACGCTCGCGCTTCAAGTTCAAGGCTTCCAGGCGGGGCGGAGATCTGCTGGGCAA 2927  
Db 1440 GCAGCAGACTCGTACAAGTTCTGGCGTTCAACGCGGGCGGAGGTGTGCTGGGCAA 1499  
QY 2928 GGAATCGGGTACCTCGATGAGATGGCGTGGCCATCTCTTTCGCTTCTTACAGCTT 2987  
Db 1500 GGAATCGGGTACCTCGATGAGATGGCGTGGCCATCTCTTTCGCTTCTTACAGCTT 1559  
QY 2988 CCGGCTGCTGGAGGGGACCCGGTGCAGTACCGCATGATGACCATCTCTTCCATGGCGCA 3047  
Db 1560 GACCGTGGCGCGGGCCACCGGTGGAGCAGAAGATGTCTGCTCAGCTCTTTCATGAAGG 1619  
QY 3048 CGGCT 3053  
Db 1620 CGGCT 1625

## RESULT 3

US-09-410-551B-1  
; Sequence 1, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER

; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; TITLE OF INVENTION: CONSTRUCTS THEREFOR  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 77536  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (52275)....(71465)  
US-09-410-551B-1

Query Match 1.7%; Score 65.4; DB 4; Length 77536;  
Best Local Similarity 51.9%; Pred. No. 1.5e-05;  
Matches 175; Conservative 0; Mismatches 156; Indels 6; Gaps 1;

QY 2232 ATCGAGTAGGCGAGCCGCGACACGCGCGCGCTTCGGGAGCAGACAGCCTCCGG 2291  
Db 56223 ACCGGCCCGCGGTGGACACCTGTCCCGCGCAGACGGGTGTTCGGCTGACCCCGGGC 56282  
QY 2292 GACGTGGTCTCAACTTCGTGATCGCGGCGGGACACACGCGCAGCAGCTGTCTGG 2351  
Db 56283 GGCATCGGCCCGAGCGGCTTACCGACCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 56342  
QY 2352 TTCACGCACATGGCCATGTCCACCCGAGCGTGGCGGAGAGCTGCGCGCGAGCTGTGC 2411  
Db 56343 AGCTTACACACGCGCGCTCCGTCCGATCGTGTTCGGACCGCGTGTGACGGCTGTGC 56402  
QY 2412 GGTTCGAGGCGGCGCGCGCGCGAGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2471  
Db 56403 GACCTCGGCACACTGCGCGCGCGAGAGTCTCTGCTCACGCGGCCACCGCGCGGTGTC 56462  
QY 2472 GACGCGCAGCAGAGGCGTTCGCGCGCGCGCGCGAGTTCGCGGGGCTCTCTCACCTAC 2531  
Db 56463 GGCATGGCGCGCGCACAGATCGCGCGCACCTGG-----GGCGGAGCTCTACGCCACC 56516  
QY 2532 GACGCTCGGCAAGCTGTCTTACCTCAACGCTGCG 2568  
Db 56517 GCAGTAGCGGCAAGCAGCAGCTCTCTGCGCGCGCGCG 56553

## RESULT 4

US-09-452-239-41  
; Sequence 41, Application US/09452239  
; Patent No. 6465229  
; GENERAL INFORMATION:  
; APPLICANT: Rafalski, Antoni J.  
; APPLICANT: Fader, Gary M.  
; APPLICANT: Cahoon, Rebecca E.  
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase  
; FILE REFERENCE: BB1284 US NA  
; CURRENT APPLICATION NUMBER: US/09/452,239  
; CURRENT FILING DATE: 1999-12-01  
; EARLIER APPLICATION NUMBER: 60/110,594  
; EARLIER FILING DATE: 1998-December-02  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 41  
; LENGTH: 1078



```

; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-452-239-41

Query Match      1.6%; Score 62.8; DB 4; Length 1078;
Best Local Similarity 48.1%; Pred. No. 5.9e-06;
Matches 178; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 2252 CGACGACGCGCGCTTCGGGAGACGACAGAGCCTCCGGGACGTGGTCTCAACTTCGT 2311
DB 423 CCCCACCGGCTCGCCATCCCGACGACGCGACCATCTTGGCCATGACATCAACCGGGA 482
QY 2312 GATCCCGGGGGGACACGACGCGGACGACGCTGCTGTTTACGACATGCCCATGTC 2371
DB 483 GAACTACGAGCTGGGGCTGCGGTGATCGAGAGCGCGGCTGGGCGCAAGATCGACTT 542
QY 2372 CCACCCGAGTGGCCGAGAGCTCGGCGGAGCTGTCGCTTCGAGGCGGAGCGGCG 2431
DB 543 CGCGAGGCGCGGCGCTGCGGCTGCTGGAACGCGCTGCGGAGGAGGAGGCGCAACCA 602
QY 2432 CGCGAGGAGGCGCTGCGGCTGCTGCGGCGGCGGCTGAGCGGAGGAGCAAGGCGTT 2491
DB 603 CACTTCGACTTCGTCCTGCGGAGCGCGGACGACGACAACTACTCAACTACGAGCG 662
QY 2492 CGCGCGCGCGTGGCGGAGTTCGGGGGCTCTCACTACGACAGCCTCGGCAAGCTGGT 2551
DB 663 CTTATGAGGCTGCTCAAGCTCGCGGCTCTCGGCTACGACACACGCTCTGGAACGG 722
QY 2552 CTACTTCACGCTGCTCAGGACGCTCGGCTGCTACCGCGGCTGCTGCTGAGTGG 2611
DB 723 CTCGCTGCTGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 782
QY 2612 CGCGCGCGGAC 2621
DB 783 GTCCTCGAC 792

RESULT 5
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match      1.5%; Score 59.8; DB 4; Length 1926;
Best Local Similarity 43.9%; Pred. No. 4.7e-05;
Matches 306; Conservative 0; Mismatches 387; Indels 4; Gaps 1;

QY 2241 GCGGAGCGCGGACGACGCGCGGCTTCGGGAGACGACGAGCCTCCGGGACGTGGT 2300
DB 952 GACGGGAGGAGGACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGAG 893
QY 2301 CTCAACTTCGTATCGCGGCGGAGACGACGCGGACGCGCTGCTGCTGCTTCAAGC 2360
DB 892 GACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGAGGAGGAGGAGGAC 833
QY 2361 ATGGCCATGTCCTCCCGGAGCGTGGCGGAGGAGCTGCGCGCGGAGCTGTGCGGTTCCGAG 2420

; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-452-239-41

Query Match      1.6%; Score 62.8; DB 4; Length 1078;
Best Local Similarity 48.1%; Pred. No. 5.9e-06;
Matches 178; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 2252 CGACGACGCGCGCTTCGGGAGACGACAGAGCCTCCGGGACGTGGTCTCAACTTCGT 2311
DB 423 CCCCACCGGCTCGCCATCCCGACGACGCGACCATCTTGGCCATGACATCAACCGGGA 482
QY 2312 GATCCCGGGGGGACACGACGCGGACGACGCTGCTGTTTACGACATGCCCATGTC 2371
DB 483 GAACTACGAGCTGGGGCTGCGGTGATCGAGAGCGCGGCTGGGCGCAAGATCGACTT 542
QY 2372 CCACCCGAGTGGCCGAGAGCTCGGCGGAGCTGTCGCTTCGAGGCGGAGCGGCG 2431
DB 543 CGCGAGGCGCGGCGCTGCGGCTGCTGGAACGCGCTGCGGAGGAGGAGGCGCAACCA 602
QY 2432 CGCGAGGAGGCGCTGCGGCTGCTGCGGCGGCGGCTGAGCGGAGGAGCAAGGCGTT 2491
DB 603 CACTTCGACTTCGTCCTGCGGAGCGCGGACGACGACAACTACTCAACTACGAGCG 662
QY 2492 CGCGCGCGCGTGGCGGAGTTCGGGGGCTCTCACTACGACAGCCTCGGCAAGCTGGT 2551
DB 663 CTTATGAGGCTGCTCAAGCTCGCGGCTCTCGGCTACGACACACGCTCTGGAACGG 722
QY 2552 CTACTTCACGCTGCTCAGGACGCTCGGCTGCTACCGCGGCTGCTGCTGAGTGG 2611
DB 723 CTCGCTGCTGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 782
QY 2612 CGCGCGCGGAC 2621
DB 783 GTCCTCGAC 792

RESULT 6
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5978807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match      1.5%; Score 59.8; DB 2; Length 1931;
Best Local Similarity 43.9%; Pred. No. 4.7e-05;
Matches 306; Conservative 0; Mismatches 387; Indels 4; Gaps 1;

QY 2241 GCGGAGCGCGGACGACGCGCGGCTTCGGGAGGACGACGAGCCTCCGGGACGTGGT 2300
DB 952 GACGGGAGGAGGACGAGGACGCGGAGGAGGACGCGGAGGAGGAGGAGGAGGAG 893
QY 2301 CTCAACTTCGTATCGCGGCGGAGACGACGCGGACGCGCTGCTGCTGCTTCAAGC 2360
DB 892 GACGAGGACGCGGAGGAGGACGCGGAGGAGGACGAGGACGCGGAGGAGGAGGAGGAC 833
QY 2361 ATGGCCATGTCCTCCCGGAGCGTGGCGGAGGAGCTGCGCGCGGAGCTGTGCGGTTCCGAG 2420
```



LOCATION: 46851...47891  
 OTHER INFORMATION: /product= "Sorm"  
 OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the polyketide rapamycin."  
 US-08-764-233A-1

Query Match  
 Best Local Similarity 1.5%; Score 58.2; DB 1; Length 49377;  
 Matches 141; Conservative 0; Mismatches 138; Indels 0; Gaps 0;  
 QY 2243 CGAGCCGGCGACGACGCGCGCGCTTCGGGGAGCAAGAGCTTCGGGACGCTCGGGACGCTGCT 2302  
 DB 8502 CGAGCCGGCGATGGCGGGGCTCTTCGAGGATGACGTGAGCTTGGTCCGGAGATC 8443  
 QY 2303 CAATCTCGTATCCCGCGCGGACACGACGCGGACGACGCTGCTGCTTCAGCCAT 2362  
 DB 8442 CGAAGCGGAGACCGCGCGCGGGGACGCCATCGCTGCGGACCCCAAGCCCTCG 8383  
 QY 2363 GGCCATGTCCACCCGACGCTGGCGGAGAGCTGCGCGGAGCTGCGGGTTCGAGGC 2422  
 DB 8382 TGGCATCGACGCGGACGCGGAGCGCATCCACGCGATCAGGGGATTCGCTGGGGTG 8323  
 QY 2423 GGAGCGCGCGGAGAGGCGCTGCGCTGCTGCGCGCGCTGACGCGGACGA 2482  
 DB 8322 TGTGAGGCTGGGGGCGAGGCGCTTGTGAGGAGGAGGACGATCTTGGCGAGCCCG 8263  
 QY 2483 CAAGGCGTTCGCGCGCGCTGGCGGAGTTCGCGGGCT 2521  
 DB 8262 CGAGCGCGCGCGGCGCTCGAGGTGGCCCAACGTTGGTCT 8224

## RESULT 8

US-09-197-649-7  
 ; Sequence 7, Application US/09197649  
 ; Patent No. 6194550  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gold, Larry  
 ; APPLICANT: Turk, Craig  
 ; APPLICANT: Pribnow, David  
 ; APPLICANT: Smith, Jonathan D.  
 ; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation  
 ; FILE REFERENCE: NEX02/C1-CON  
 ; CURRENT APPLICATION NUMBER: US/09/197,649  
 ; CURRENT FILING DATE: 1998-11-23  
 ; EARLIER FILING DATE: 1998-11-23  
 ; EARLIER APPLICATION NUMBER: 07/829,461  
 ; EARLIER FILING DATE: 1992-01-31  
 ; EARLIER APPLICATION NUMBER: 07/739,055  
 ; EARLIER FILING DATE: 1991-08-01  
 ; EARLIER APPLICATION NUMBER: 07/561,968  
 ; EARLIER FILING DATE: 1990-08-02  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 7  
 ; LENGTH: 390  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Sequence  
 ; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed  
 ; OTHER INFORMATION: fragments having NcoI restriction sites.  
 US-09-197-649-7

Query Match  
 Best Local Similarity 1.5%; Score 58; DB 3; Length 390;  
 Matches 195; Conservative 0; Mismatches 190; Indels 3; Gaps 2;  
 QY 2227 GGTTCATCAGCTAGCGGAGCGCGGACGACGCGCGGCTTCGGGACGACAGCC 2286  
 DB 1 GGGCCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
 QY 2287 TCCGGACGTGTGTCACTTCGTGATCGCGCGCGGGAC-ACGACGCGACGACGCTG 2345

DB 61 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 120  
 QY 2346 TCGTGGTTACGACATGGCCATCTC--CCACCCGGAGGTGGCGAGAGCTGCGCCCG 2403  
 DB 121 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 180  
 QY 2404 AGCTGTGCGCTTCGAGGCGGAGCGCGCGGCGGAGGCGCTCGCTCGTGCCTGCG 2463  
 DB 181 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 240  
 QY 2464 GCGCGCTGAGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2523  
 DB 241 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 300  
 QY 2524 TCACCTACGACGACGCTCGGCAAGCTGCTTACCTCCAGGCTGCTCACCAGAGCGTCC 2583  
 DB 301 ACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 360  
 QY 2584 GCCTGTACCCCGCGCTCCCTCAGGTGAG 2611  
 DB 361 ACGACGACGACGATGTTGCGATGTCAG 388

## RESULT 9

US-09-266-965-98  
 ; Sequence 98, Application US/09266965  
 ; Patent No. 6495348  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D  
 ; APPLICANT: Mao, Y  
 ; APPLICANT: Varoglu, M  
 ; APPLICANT: He, M  
 ; APPLICANT: Sheldon, P  
 ; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
 ; FILE REFERENCE: 600.456US1  
 ; CURRENT APPLICATION NUMBER: US/09/266,965  
 ; CURRENT FILING DATE: 1999-03-12  
 ; EARLIER FILING DATE: 1999-03-12  
 ; EARLIER FILING DATE: 1996-08-19  
 ; EARLIER APPLICATION NUMBER: PCT/US94/11279  
 ; EARLIER FILING DATE: 1994-10-06  
 ; EARLIER APPLICATION NUMBER: US 08/133,963  
 ; EARLIER FILING DATE: 1993-10-07  
 ; NUMBER OF SEQ ID NOS: 145  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 98  
 ; LENGTH: 819  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces lavendulae  
 US-09-266-965-98

Query Match  
 Best Local Similarity 1.5%; Score 57.4; DB 4; Length 819;  
 Matches 255; Conservative 0; Mismatches 291; Indels 3; Gaps 2;  
 QY 2329 CGACGCGACGACGCTGCTGCTTACGACATGCGCATGTCACCCGCGGACGCTGCGCG 2388  
 DB 6 CCGTCTCGACGACGACGACGCGGGGCGACGCGGTGATCACTGATCGCGCTGACAGGCGA 65  
 QY 2389 AGAAGCTGGCGCGAGCTGTCGCGCTTCGAGGCGGACGCGCGCGGAGGAGGCGCTCG 2448  
 DB 66 GATGAGCGGCTTCTTCCCGCTGCGGTGCGGGACCGC-TGCGGGGCGGCGCGCG 124  
 QY 2449 CGCTGCTCCCTGCGCGCGCTGACGCGGACGACGACGCTTCCCGCGCGCGCTGCGCG 2508  
 DB 125 CCGGACGCGCGCGCGCGCGCGCTGCGCGCGCGCTGCTCTCTCTCGACGACGACG 184  
 QY 2509 AGTTGCGCGGCTCTCTACCTACGACGCTCGGCAAGCTGCTTACCTCCACGCTCGG 2568  
 DB 185 TCCTGGTGGCGCGGCTTCTTCCCGCACGCGCGCGCGCGGACCGGACGCGCTTCA 244  
 QY 2569 TCACGAGACGCTCGCGCTGTACCGCGCGCTTCTCTAGGTGAGCGCGCGCGACGCG--G 2626

Db 245 CCACGCGCGCTCGCGGAAGTCCCGACCGCGCGCGGTTCTCGCGCTGTGAGAAGG 304  
QY 2627 ACCTCGGTCGAGACACAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2686  
Db 305 CGCCCGGACCGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 364  
QY 2687 CGCGCGCGGAGGACCCCAAGGGGATCTGAGGACGACGAGTGTGCGGACGGAAGG 2746  
Db 365 CCAGCGCGCGCAACCGACCGCGCGCTCGTCCCAACCGCTTGGAGCGCGCGCGTGGAG 424  
QY 2747 TGAGGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2806  
Db 425 CCATGCGCGCGCGCTCCCTGCGGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 484  
QY 2807 GGGGCGCGGACCGCGGAGTTCGCGCGCGGAGTTCGCGCGCGGAGTTCGCGCGCG 2866  
Db 485 CGCGCGCGGACGAGCGCGGATGGAGCATACCGCGCGGATTCGAGGAGGAGTTCGGGCTCA 544  
QY 2867 GCAAGCGGT 2875  
Db 545 CTGGGGGT 553

## RESULT 10

US-09-452-239-35

; Sequence 35, Application US/09452239

; Patent No. 6465229

; GENERAL INFORMATION:

; APPLICANT: Rafalski, Antoni J.

; APPLICANT: Fader, Gary M.

; APPLICANT: Cahoon, Rebecca E.

; TITLE OF INVENTION: Plant Caffeyol-CoA O-Methyltransferase

; FILE REFERENCE: B01284 US NA

; CURRENT APPLICATION NUMBER: US/09/452,239

; CURRENT FILING DATE: 1999-12-01

; EARLIER APPLICATION NUMBER: 60/110,594

; EARLIER FILING DATE: 1998-December-02

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 35

; LENGTH: 1018

; TYPE: DNA

; ORGANISM: Triticum aestivum

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (817)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (826)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (874)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (891)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (924)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (934)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (961)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (970) .. (971)

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1012)

; US-09-452-239-35

Query Match 1.5%; Score 57.4; DB 4; Length 1018;  
Best Local Similarity 49.2%; Pred. No. 0.00013;  
Matches 151; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 2252 CGACGACGCGCGCGCTCGGGGACGACAAAGAGCTTCCGGGACGCTGCTCAACTTGGT 2311  
Db 400 CGCCACCGCGCTCGGCATCCCGACGACGCGCACCATTTGGCCATGGACATCAACCGCA 459

QY 2312 GATCCCGCGCGGACACGACGCGACGACGCTGTGTGTGTTCAGGACATGGCCATGTC 2371  
Db 460 GAACTACGAGCTGGGGCTGCCGTGCATCGAAGGCGCGCTGGCGGCAAAAGATCGACTT 519

QY 2372 CCACCGGACGCTGGCGGAGAGCTCGCGCGGAGCTGTGGGTTTCGAGGCGGAGCGGC 2431  
Db 520 CGCGGAGCGCGCGGCTGCCGTCTGACGCGCTGTGGAGGACGAGGCGCAACCGG 579

QY 2432 GCGGAGGAGGCGCTGCGGCTGTGTCCTTGGCGCGCGCTGACGCGGACGACAAAGCGTT 2491  
Db 580 GACCTTCGACTTGGTCTTCTGTTGGACGCGGACAAAGGACAACTACCTCACTACGAGG 639

QY 2492 CGCGCGCGCGTGGCGGAGTTGCGGGGCTCTTACCTACGACAGCTTGGCAAGCTGGT 2551  
Db 640 CTTATGAAGCTCGTCAAGGTGCGCGGCTCTCTCGGCTACGACAAACCCCTCTGGAACG 699

QY 2552 CTACCTC 2558  
Db 700 CTCGCTC 706

## RESULT 11

US-09-266-965-9

; Sequence 9, Application US/09266965

; Patent No. 6495348

; GENERAL INFORMATION:

; APPLICANT: Sherman, D

; APPLICANT: Mao, Y

; APPLICANT: Varoglu, M

; APPLICANT: He, M

; APPLICANT: Sheldon, P

; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster

; FILE REFERENCE: 600.456US1

; CURRENT APPLICATION NUMBER: US/09/266,965

; CURRENT FILING DATE: 1999-03-12

; EARLIER APPLICATION NUMBER: US 08/624,447

; EARLIER FILING DATE: 1996-08-19

; EARLIER APPLICATION NUMBER: PCT/US94/11279

; EARLIER FILING DATE: 1994-10-06

; EARLIER APPLICATION NUMBER: US 08/133,963

; EARLIER FILING DATE: 1993-10-07

; NUMBER OF SEQ ID NOS: 145

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 9

; LENGTH: 3765

; TYPE: DNA

; ORGANISM: Streptomyces lavendulae

; US-09-266-965-9

Query Match 1.5%; Score 57.4; DB 4; Length 3765;  
Best Local Similarity 46.4%; Pred. No. 0.00028;  
Matches 255; Conservative 0; Mismatches 291; Indels 3; Gaps 2;

QY 2329 CGACGCGGACGACGCTGTGCTGTTTACGACATGCCATGTCCTCCACCGGAGTGGCGC 2388  
Db 1884 CGTCGTCGACGACAGCGGGGCGACGCCGCTGATCACTGATCGCGCTGACAGCGCA 1943

QY 2389 AGAAGCTGCGCGGAGTGTGCGGTTTCGAGCGGAGCGCGCGCGGAGGCGGTGCG 2448  
Db 1944 GATGAGCGGCTTCTCCCGCTGCGCGTGTGCGGGGACCGG-TGCGGGGCGGGCGCG 2002

QY 2449 CGCTGTCCTTGGCGGCGGCTGACCGGACGACAAAGCGTTCCCGCGCTGGCGC 2508  
Db 2003 CCGGAACGCGGGGCGCGCGCTCGGCGCGGCTGCTGCTTCTTCGACGACGACG 2062

	Query Match	1.5%	Score 57.4	DB 4	Length 53500
	Best Local Similarity	46.4%	Pred. No. 0.0013		
	Matches 255	Conservative 0	Mismatches 291	Indels 3	Gaps 2
QY	2329	CGACGGCAGCAGCGTGTCTGTGGTTCACGCACATGGCCATGTCCACCCGACGCTGGCCG	2388		
DB	20152	CGTCTGTCCAGCACACGACGGGGGCGACCCCGGTGATCAATGATCCCGCTGACAGCGA	20093		
QY	2389	AGAACTCGCGCGCGAGCTGTTCGCGGTTTCGAGGCGGAGCGCGCGAGGAGGGCTCG	2448		
DB	20092	GATGAGGGCTTCTCCGCTGTCCGCTGTGTGCGGGGACCGC-TGCGGGGCGGGCGCGCG	20034		
QY	2449	CGCTGTGCTCCTCGCGGGCGCTGACGCGACGCAAGGGCTTTCGCGCCCGCTGGCGC	2508		
DB	20033	CCCGAAACGCGGGGCGCGCGCCCTCGGCCCGCCGCTGTCTCTCGACGACGACG	19974		

	Query Match	1.5%;	Score 57.2;	DB 4;	Length 3794;
	Best Local Similarity	49.7%;	Pred. No. 0.00032;		
	Matches 146;	Conservative 0;	Mismatches 148;	Indels 0;	Gaps 0;
2321	QY	CGCGGACACACGCGACGAGCTGTCGTTGTCACGCATGTCCTCCACCCGGA	2380		
2869	Db	GCTGTCCAAACAGGAGTTCGGCGGGCGGTCTGGCGCGGAAGACGACGGCTCGAGGA	2928		
2381	QY	CGTGGCCGACAAGCTTCGGCCGCGAGCTGTGCGGTTGAGGCGGAGCGCGCGCGGAGGA	2440		
2329	Db	GCTGTGTGGCGAGGTGGCCCGGCTGTGAGGCGGAGAAACGGCGACGACGAGGCCCGCACCGC	2988		
2441	QY	GGGCGTGCGCCTCGTGTCCCTTGCGCGCGGCGCTCAGCGCGACGACAGGGGTTGCGCGCCG	2500		
2989	Db	GGGCGCTGGAGCGGACCTTGGCAGGGTGGACGGCGACAACGGCGTCTGTGCGCGCCCGCCA	3048		
2501	QY	CGTGGCGCAGTTGCGGGGCGCTCTCACCTACGACAGCTTCGGCAGCTGGTCTACCTCCA	2560		
3049	Db	CGCGAGCTGGCGCGCGCCCTCAGTCGCTGGGGCGGCGTCTCTCGAGGTGTCGAGATGCG	3108		

[illegible]

```

RESULT 14
US-08-682-847-1
; Sequence 1, Application US/08682847
; Patent No. 5858989
; GENERAL INFORMATION:
; APPLICANT: BABIUK, LORNE
; APPLICANT: VAN DEN HURK, SYLVIA
; APPLICANT: ZAMB, TIM
; APPLICANT: FITZPATRICK, DAVID
; TITLE OF INVENTION: RECOMBINANT BOVINE HERPESVIRUS TYPE 1
; TITLE OF INVENTION: POLYPEPTIDES AND VACCINES

```

```

/ CORRESPONDENCE ADDRESS:
/
/ ADDRESSEE: MORRISON & FOERSTER
/ STREET: 755 PAGE MILL ROAD
/ CITY: PALO ALTO
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304-1018
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/

```

APPLICATION NUMBER: US/08/682,847  
FILING DATE: 12-JUL-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:

NAME: PARK, FREDDIE K.  
REGISTRATION NUMBER: 35,636  
REFERENCE/DOCKET NUMBER: 29310-20005.10  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEC ID NO. 1

SEQUENCE CHARACTERISTICS:  
LENGTH: 3382 base pairs  
TYPE: nucleic acid  
STRANDNESS: 1

STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:

NAME/KEY: CDS  
LOCATION: 432..3230  
S-08-682-847-1

```

Query Match          1.4%; Score 56.4; DB 2; Length 3382;
Best Local Similarity 51.6%; Pred. No. 0.00048;
Matches 129; Conservative 0; Mismatches 121; Indels 0

```

2297	GGTGCTCAACTTCGTGATGCGCGCGCGGACACGACGCGCTGTCGTGTTTAC	2356
1829	GATGCTCAGCAGAGCTGGCCACAGCTGACCTGAGGAGCTGGCGGCTCGAAGCGAC	1888

[illegible][illegible][illegible]

2002 CCCCCCGGGGAGACATCCGGTAGCTCGGCCGAGTTTGGCGCGGTGCAGTTCACCTACGA 2068

QY 2537 CCTCGGCAG 2546  
||| ||| |||  
DB 2069 CCACTTCCT 2078

RESULT 15  
PCT-US95-04801-3  
; Sequence 3, Application PC/TUS9504801  
; GENERAL INFORMATION:  
; APPLICANT: Martin, Juan F.  
; APPLICANT: Coque, Juan R.  
; APPLICANT: Esquita, Francisco J.  
; APPLICANT: Fuente, Juan L.  
; APPLICANT: Liarena, Francisco J.  
; APPLICANT: Liras, Paloma

/ / TITLE OF INVENTION: LATE GENES  
 / /  
 / / NUMBER OF SEQUENCES: 8  
 / / CORRESPONDENCE ADDRESSES:  
 / / ADDRESS: John W. Wallen III  
 / / STREET: P.O. Box 2000  
 / / CITY: Rahway  
 / / STATE: New Jersey  
 / / COUNTRY: USA  
 / / ZIP: 07065  
 / /  
 / / COMPUTER READABLE FORM:  
 / / MEDIUM TYPE: Floppy disk  
 / / COMPUTER: IBM PC compatible  
 / / OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04801
; FILING DATE:

```

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wallen III, John W.  
REGISTRATION NUMBER: 35 403

REFERENCE/DOCKET NUMBER: 19179  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3905  
TELEFAX: (908) 594-4720

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 900 base pairs  
TYPE: nucleic acid

NAME: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA

Query Match 1.43; Score 56; DB 5; Length 900;  
Best Local Similarity 44.33; Pred. No. 0.00028;

best local similarity 44.3%; Pred. No. 0.00028;  
 Matches 271; Conservative 0; Mismatches 340; Indels 1; Gaps 1;  
 2244 GAGCGCGCGACACGCGCGCGCGCTTCGGGAGCGACAAGAGCGCTCGGAGCGCTGGCTC 2103

50 GACGGCGGGACGGGACTGTGTCATCGACGCCGTGACCCGGCCGCGGAGGTGTCTTC 109

2304 AACTTCGTGATCCGCCGGGGCGGACACGACGGCGGACGACGCGTGTCTGTGGTTTACGGCACATG 2363

110 AACTTCGGCAAGGTGGGCGCTGGAGACGACCATCACCGACTTGGCGAGGCTCGATCCGG 169

2364 GCCATGTCCTCCACCCCGACGTCGGCCGAGAGCTGCGCGCGAGAGTCTGCGCGGCTTCAACCCG 2422

[illegible]

b

230 GGCGTCTGGACAGCGGAAACCCGGGTGGAGCAGTACCGCGGGAAAACCGTGACTG 289

2484 AACGCGTCATCCC

2181 ATGGGGGATCGCCGCTCCGCGGCTCTCAGCTACGACAGCTCTCGGC 2543

```
Db      290 CTCCTGCTCCTCACCCTGCGCGGACGTGGTGGAGTTCTTCGACGCCACCCTGCGGGCGCAG 349
QY      2544 AAGCTGGTCTTACCTCCACGCTGCTGTCACCGAGACGCTCGGCTGTACCCCGCGCTCCCT 2603
Db      350 GACGCGCGCGACGACCCGCGCGCCAGTCCCGCACCCAGCGGTGCACGTGACACGAGC 409
QY      2604 CAGGTGAGCGCGCGCGACACGCGACCTCCGGTCCAGAGCACAGCATGTGAGTGGACC 2663
Db      410 CCGGGCAGCGCGCGCGCCACAGGCGCGACCTCGGCCCCCGCGGAGTTCGGGCGC 469
QY      2664 TGAATGCAATGCACATGCACTTGGCGCGCGCGAGACCCCAAGGGATCCTGGAGGACGA 2723
Db      470 TTCCAGATCATCAACGTCTGGCGCGCGC-TGCTCGAGCGGTGCGCAACTTCCCGCTGSC 528
QY      2724 CGTGTGCGCGACGGGACGAAGGTGAGGCGCGCGGATGGTGACGTAGTGGCCCTACTC 2783
Db      529 GCTGTGCGACTACCGGTGCTGGACCTGTCCGCCACCTGGTGGCGACCCCGCTGGAATT 588
QY      2784 GATGGGCGGATGGAGTCAAACTGGGGCCCCGACCGCGCGAGCTTCCGCGCGGAGCGGTG 2843
Db      589 CCCGGACTGGGTGAAGGACCGGAGAACTACTCGGTCCGGCACACCCCGCGCACCGCTG 648
QY      2844 GATCAACGAGGA 2855
Db      649 GTACTTCGGGA 660
```

Search completed: March 7, 2004, 17:06:42  
Job time : 195 secs



us-10-021-657-7.rng

Mon Mar 8 10:20:24 2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2004, 10:14:42 ; Search time 961 Seconds  
(without alignments)  
17227.084 Million cell updates/sec

Title: US-10-021-657-7  
Perfect score: 1 gaattccaggagccctt.....gtcgagtaaacgaggaattc 3897

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues  
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N Geneseq 29Jan04 : \*  
1: Geneseqn1980s : \*  
2: Geneseqn1990s : \*  
3: Geneseqn2000s : \*  
4: Geneseqn2001as : \*  
5: Geneseqn2001bs : \*  
6: Geneseqn2002as : \*  
7: Geneseqn2003as : \*  
8: Geneseqn2003bs : \*  
9: Geneseqn2003cs : \*  
10: Geneseqn2004s : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3897	100.0	3897	6	ABL60615 Maize SBM
2	1378.4	35.4	1906	6	ABL60611 Maize SBM
3	1092	28.0	1092	6	ABL60612 Sorghum S
4	268.4	6.9	494	6	ABL60614 Maize SBM
5	263	6.7	267	6	ADA70596 Rice gene
6	137.8	3.5	1527	7	ADA71342 Rice gene
7	127	3.3	1536	6	ACL21310 DNA clone
8	120.8	3.1	657	8	ACL21310 Codon mod
9	119.6	3.1	1626	2	AAX26878 Cytochrom
10	119.6	3.1	2181	2	AAX26874 Rice gene
11	114	2.9	1563	7	ADA70663 DNA clone
12	112.8	2.9	621	8	ADA49310 Maize gene
13	105.4	2.7	528	8	ACL21305 DNA clone
14	104	2.7	399	8	ACL21309 DNA clone
15	104	2.7	507	8	ACL21308 DNA clone
16	104	2.7	527	8	ACL21304 DNA clone
17	104	2.7	1652	6	ABZ13076 Arabidops
18	104	2.6	1965	7	ADA71125 Rice gene
19	100.6	2.6	1506	7	ADA70631 Rice gene
20	99.4	2.6	1614	6	ABZ14873 Arabidops
21	96	2.5	1896	3	AAC45033 Arabidops
22	96	2.5	1896	3	AAC45033 Arabidops
23	93.8	2.4	866	6	ABN98804 Arabidops

24	93.2	2.4	1395	8	ADA48303 Rice gene
25	92.4	2.4	1542	6	ABZ14063 Arabidops
26	92.4	2.4	1736	2	AAX58407 DNA seque
27	90	2.3	1473	6	ABZ14176 Arabidops
28	86	2.2	843	8	ADA48919 Wheat gen
29	81.6	2.1	1694	3	AAC37317 Arabidops
30	81.4	2.1	1488	6	ABZ13809 Arabidops
31	81.4	2.1	1488	7	ADA68342 Arabidops
32	79	2.0	1440	3	AAC42927 Arabidops
33	78.6	2.0	589	3	AAC38695 Arabidops
34	78.4	2.0	1117	3	AAC40695 Arabidops
35	78.4	2.0	1533	3	AAC42704 Banana ge
36	78.2	2.0	1537	8	ADA48779 Arabidops
37	75.2	1.9	1551	3	AAC43122 Arabidops
38	74.4	1.9	1407	3	AAC33115 Arabidops
39	74	1.9	678	8	ACL21311 DNA clone
40	72.2	1.8	31263	7	ACA37577 Prokaryot
41	70.8	1.8	818	4	AHH06808 Human CDN
42	68.6	1.8	1961	4	AHH14525 Human sec
43	68.6	1.8	3871	2	AAX50801 Human sec
44	68.2	1.8	3871	4	AAS9207 Human.CDN
45	68.2	1.8	3871	4	AAS9207 Human.CDN

ALIGNMENTS

RESULT 1  
ABL60615 standard; DNA; 3897 BP.  
ID ABL60615 standard; DNA; 3897 BP.  
XX  
AC ABL60615;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Maize SBMu200 genomic DNA sequence.  
XX  
KW SBMu200; fertility; maize; hybrid seed; sterility; plant; gene;  
XX  
KW transgenic; ds.  
XX  
OS Zea mays.  
XX  
PN WO200226789-A2.  
XX  
PD 04-APR-2002.  
XX  
PF 25-SEP-2001; 2001WO-US029886.  
XX  
PR 26-SEP-2000; 2000US-00670153.  
XX  
XX (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Albertsen MC, Fox T, Huffman G, Trimmell M;  
XX  
XX WPI; 2002-471258/50.  
XX  
XX Novel nucleic acid sequence comprising SBMu200 gene, useful for producing  
PT transgenic plants and mediating male fertility in plants.  
XX  
XX Claim 3; Fig 5; 60pp; English.  
XX  
PS The invention relates to a SBMu200 gene, useful for impacting male  
XX fertility in a plant e.g. maize. An isolated nucleotide sequence  
XX comprising the SBMu200 gene is useful for producing hybrid seeds and for  
CC providing heritable male sterility in a plant.  
CC The present sequence represents the maize SBMu200 genomic DNA sequence  
CC  
XX  
SQ Sequence 3897 BP; 927 A; 1013 C; 1058 G; 899 T; 0 U; 0 Other;

Query Match 100.0%; Score 3897; DB 6; Length 3897;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3897; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1	GAATTC	CAAGCGAGG	CCCTT	GTAG	CAGAGAGT	GTGCT	GATG	CAGT	CGCGG	AAATG	AGT	60
1	GAATTC	CAAGCGAGG	CCCTT	GTAG	CAGAGAGT	GTGCT	GATG	CAGT	CGCGG	AAATG	AGT	60
61	GGGTCT	GAGAGC	ACGCT	GAGGG	GTCC	AGGAT	GCAAT	GGCT	ATG	GCAAT	CGGCT	120
61	GGGTCT	GAGAGC	ACGCT	GAGGG	GTCC	AGGAT	GCAAT	GGCT	ATG	GCAAT	CGGCT	120
121	AGGTG	GAGGAC	AAAGT	GTG	AGAGT	TTGG	GAGG	GCAAC	CTAT	GGAAGT	TTGGT	180
121	AGGTG	GAGGAC	AAAGT	GTG	AGAGT	TTGG	GAGG	GCAAC	CTAT	GGAAGT	TTGGT	180
181	AGC	CAATG	AGAGAT	CTAT	TC	CAGACT	TAC	CTG	GAATG	CGCC	CAACAAAT	240
181	AGC	CAATG	AGAGAT	CTAT	TC	CAGACT	TAC	CTG	GAATG	CGCC	CAACAAAT	240
241	TTT	TGAT	CTG	TAC	CTT	ACCT	TAT	TCT	TTG	TTG	TTG	300
241	TTT	TGAT	CTG	TAC	CTT	ACCT	TAT	TCT	TTG	TTG	TTG	300
301	AAT	CAAT	GAT	TAG	AT	TAG	CAAAAT	TAT	CTT	TTG	TTG	360
301	AAT	CAAT	GAT	TAG	AT	TAG	CAAAAT	TAT	CTT	TTG	TTG	360
361	GGT	GAG	ACG	AT	TAAAT	AT	CAAT	CAAT	AGAG	CTT	AT	420
361	GGT	GAG	ACG	AT	TAAAT	AT	CAAT	CAAT	AGAG	CTT	AT	420
421	TTT	CAG	AT	CAAT	CTT	CAG	TTT	CA	AGAA	TTT	CT	480
421	TTT	CAG	AT	CAAT	CTT	CAG	TTT	CA	AGAA	TTT	CT	480
481	AAG	TAG	AT	CTT	CAAT	TTT	CA	AGAA	TTT	CT	AT	540
481	AAG	TAG	AT	CTT	CAAT	TTT	CA	AGAA	TTT	CT	AT	540
541	CTT	TT	CACT	TTG	GGT	TT	CA	AAAT	GACT	CA	CAAGAA	600
541	CTT	TT	CACT	TTG	GGT	TT	CA	AAAT	GACT	CA	CAAGAA	600
601	AAT	CT	CT	CT	CAG	AT	GT	AT	CT	TT	CA	660
601	AAT	CT	CT	CT	CAG	AT	GT	AT	CT	TT	CA	660
661	CAG	TTT	TA	AGAA	CAAT	GT	CA	AGAT	TTT	CA	TT	720
661	CAG	TTT	TA	AGAA	CAAT	GT	CA	AGAT	TTT	CA	TT	720
721	CAG	AT	TTT	GT	ACCA	AGT	GT	AT	CA	CAAT	CT	780
721	CAG	AT	TTT	GT	ACCA	AGT	GT	AT	CA	CAAT	CT	780
781	ACT	GG	CA	AT	CT	CT	TAAT	TTG	AT	TA	AAAG	840
781	ACT	GG	CA	AT	CT	CT	TAAT	TTG	AT	TA	AAAG	840
841	TTG	TT	GG	CA	AT	CT	TAAT	TTG	AT	TA	AAAG	900
841	TTG	TT	GG	CA	AT	CT	TAAT	TTG	AT	TA	AAAG	900
901	GGT	CAG	GAC	CA	AT	TAA	TAA	TAA	TAA	TAA	TAA	960
901	GGT	CAG	GAC	CA	AT	TAA	TAA	TAA	TAA	TAA	TAA	960
961	TCT	CG	AGT	TTG	GGG	GT	CC	CT	TA	AGT	TTG	1020
961	TCT	CG	AGT	TTG	GGG	GT	CC	CT	TA	AGT	TTG	1020
1021	ACCC	AAAT	CCAT	GT	CT	ACAT	AT	CA	TAT	AGCAT	CCAT	1080
1021	ACCC	AAAT	CCAT	GT	CT	ACAT	AT	CA	TAT	AGCAT	CCAT	1080
1081	AGAG	CA	CCAT	GT	AGG	AA	GCT	CA	TC	AGC	CGG	1140
1081	AGAG	CA	CCAT	GT	AGG	AA	GCT	CA	TC	AGC	CGG	1140

QY	3301	CGGATGCTCCGGGACCTTTTGGAGACCATTACTGACAGCGGTGTGAATAAAAGGCTTC	3360
DB	3301	CGGATGCTTTCCGGGACCTTTTGGAGACCATTACTGACAGCGGTGTGAATAAAAGGCTTC	3360
QY	3361	TTCTCGCGCGAAGTTTTGGGTTCAAGTCTTGCGGTCTTTTCGACAGCAAAAAGGTTTGG	3420
DB	3361	TTCTCGCGCGAAGTTTTGGGTTCAAGTCTTTGGGCTTTTCGACAGCAAAAAGGTTTGG	3420
QY	3421	AAGGATCTGAACCCCTGMACCGGAAAAATGGCTTCGGAAATATGCTGCATCGGGGCGGGGCC	3480
DB	3421	AAGGATCTGAACCCCTGMACCGGAAAAATGGCTTCGGAAATATGCTGCATCGGGGCGGGGCC	3480
QY	3481	GTCACCTCGGGATGACGACAGAAGCCCCAACAGCAGTGAGAGCAAGCGATCTTTGGAGTTTGG	3540
DB	3481	GTCACCTCGGGATGACGACAGAAGCCCCAACAGCAGTGAGAGCAAGCGATCTTTGGAGTTTGG	3540
QY	3541	AGACATCTTCGHAACCCCTCGGCGCTCCGCGAGTCACTTTGCCCTCCTCTGTCGTGTCCG	3600
DB	3541	AGACATCTTCGHAACCCCTCGGCGCTCCGCGAGTCACTTTGCCCTCCTCTGTCGTGTCCG	3600
QY	3601	TGGCGGCACCGCGCCCGCCCTCGTGTTCGACCAAATCCC CGCGCCCGCACCGGTTTCTG	3660
DB	3601	TGGCGGCACCGCGCCCGCCCTCGTGTTCGACCAAATCCC CGCGCCCGCACCGGTTTCTG	3660
QY	3661	GTACAACACCCCTCATTCGCGGCGCGCGCAGTGACAGCGCCCGGAGCGCGTATACAT	3720
DB	3661	GTACAACACCCCTCATTCGCGGCGCGCGCAGTGACAGCGCCCGGAGCGCGTATACAT	3720
QY	3721	CTATAAATCATGGTATGTACTTTATTTTCAAACGGCCTTAACACCAACATATTTTATG	3780
DB	3721	CTATAAATCATGGTATGTACTTTATTTTCAAACGGCCTTAACACCAACATATTTTATG	3780
QY	3781	GTAACACAGTTTCAAAATGTACAAAATTTAAACAGGCACAAACCGTAGCTTAACATAAG	3840
DB	3781	GTAACACAGTTTCAAAATGTACAAAATTTAAACAGGCACAAACCGTAGCTTAACATAAG	3840
QY	3841	AGAATGAGAGACACCCCAAGGTTAGAGATGAATAAGCTGAGTAAACGACGAATTC	3897
DB	3841	AGAATGAGAGACACCCCAAGGTTAGAGATGAATAAGCTGAGTAAACGACGAATTC	3897

RESULT 2  
 ABL6061l standard; cDNA; 1906 BP.

XX	AC	ABL6061l;
XX	DT	27-AUG-2002 (first entry)
XX	DE	Maize SBMu200 protein encoding cDNA.
XX	KW	SBMu200; gene; fertility; maize; hybrid seed; sterility; plant;
XX	KW	transgenic; ss.
XX	OS	Zea mays.

Key	Location/Qualifiers
FT	I..1770
FT	/tag= c
FT	/product= "fused SBMu200 protein"
FT	/partial
FT	/note= "CDS encodes the N-terminal and C-terminal portion
FT	of SBMu200, separated by a stop codon. CDS lacks an
FT	initiation codon"
FT	I..1641
FT	/tag= a
FT	/product= "N-terminal portion of SBMu200"
FT	/partial
FT	/note= "CDS lacks an initiation codon"
FT	1642..1770
FT	/tag= b
FT	/product= "C-terminal portion of SBMu200"



QY 2768 CGTACGTCCTACTCGATGGCGGATGAGTACAACTGGGCCCCGACGCGCGAGCT 2827  
 Db 1328 CGTACGTCCTACTCGATGGCGGATGAGTACAACTGGGCCCCGACGCGCGAGCT 1387  
 QY 2828 TCCGGCCGAGCGGTGATCAACGAGATGGCGGTTCCGCAACGCGTCCGCTTCAAGT 2887  
 Db 1388 TCCGGCCGAGCGGTGATCAACGAGATGGCGGTTCCGCAACGCGTCCGCTTCAAGT 1447  
 QY 2888 TCACGGCGTCCAGCGCGGCGAGGATCTGCTGGGCAAGGACTCGCGCTACTCGCAGA 2947  
 Db 1448 TCACGGCGTCCAGCGCGGCGAGGATCTGCTGGGCAAGGACTCGCGCTACTCGCAGA 1507  
 QY 2948 TGAAGATGGCGCTGGCCATCTCTTGGCTTCTACAGCTTCGGCTGCTGGAGGGCACC 3007  
 Db 1508 TGAAGATGGCGCTGGCCATCTCTTGGCTTCTACAGCTTCGGCTGCTGGAGGGCACC 1567  
 QY 3008 CGGTGCGATACCGCATGATGACCATCTCTCCATGGCGCGGCTCAAGTCCGCGTCT 3067  
 Db 1568 CGGTGCGATACCGCATGATGACCATCTCTCCATGGCGCGGCTCAAGTCCGCGTCT 1627  
 QY 3068 CTAGGGCGCTGATGTCATGGCGATTG-----GGATATCATCCCGCTTAATCC----- 3117  
 Db 1628 CTAGGGCGCTGATGTCATGGCGATTG-----GGATATGATATCGTCCCGCTTAATCCAGCAG 1687  
 QY 3118 -----TTAAAAATTTGCATGCGATGTAAGGAAAGCGATGGTTCAT 3163  
 Db 1688 AATAACGCTCGTGTACAAATTTGCATGCGATGTAAGGAAAGCGATGGTTCAT 1747  
 QY 3164 TGGTGGCTTGGCTTAAAGCTTAAAGCTCCGTCGGGCTTTCGCAACACCATCAGTAG 3223  
 Db 1748 TGGTGGCTTGGCTTAAAGCTTAAAGCTCCGTCGGGCTTTCGCAACACCATCAGTAG 1807  
 QY 3224 TGTTTTGACTCTACTCTCAGTGAAGTGTAGTGACAGCATACAAGTTTCATCATATA 3283  
 Db 1808 TGTTTTGACTCTACTCTCAGTGAAGTGTAGTGACAGCATACAAGTTTCATCATATA 1867  
 QY 3284 TTATCCTCTTCTT 3297  
 Db 1868 TTATCCTCTTCTT 1881

RESULT 3  
 ABL60613  
 ID ABL60613 standard; DNA; 1092 BP.  
 AC ABL60613;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE Maize SBMu200 promoter sequence.  
 XX  
 KW SBMu200; gene; fertility; maize; hybrid seed; sterility; plant; promoter;  
 transgenic; ds.  
 XX  
 OS Zea mays.  
 XX  
 FN WO200226789-A2.  
 XX  
 PD 04-APR-2002.  
 XX  
 PF 25-SEP-2001; 2001WO-US029886.  
 XX  
 PR 26-SEP-2000; 2000US-00670153.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 XX Albertsen MC, Fox T, Huffman G, Trimmell M;  
 XX  
 DR WPI; 2002-471258/50.  
 XX  
 PT Novel nucleic acid sequence comprising SBMu200 gene, useful for producing  
 transgenic plants and mediating male fertility in plants.

XX Claim 36; Fig 8; 60pp; English.  
 XX The invention relates to a SBMu200 gene, useful for impacting male  
 CC fertility in a plant e.g. maize. An isolated nucleotide sequence  
 CC comprising the SBMu200 gene is useful for producing hybrid seeds and for  
 CC providing heritable externally controllable male sterility in a plant.  
 CC The present sequence represents the maize SBMu200 promoter sequence  
 XX

SQ Sequence 1092 BP; 313 A; 208 C; 236 G; 335 T; 0 J; 0 Other;  
 Query Match 28.0%; Score 1092; DB 6; Length 1092;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-234;  
 Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGAAGCGAGGCGCTTGTAGCAGAGAGTGTGTGTATGAGTGGCGGAAATGAGT 60  
 Db 1 GAATTCGAAGCGAGGCGCTTGTAGCAGAGAGTGTGTGTATGAGTGGCGGAAATGAGT 60  
 QY 61 GCGTGTGAGAGCAACGCTGAGGGGTTCCAGGATGCAATGGCTATGGCAATCGGCTAG 120  
 Db 61 GCGTGTGAGAGCAACGCTGAGGGGTTCCAGGATGCAATGGCTATGGCAATCGGCTAG 120  
 QY 121 AGGTGGAGGACAAAGGTGCTGAGGATTTGGGAGGCGAACCTATGGCAAGTTGGTGAAGAGGC 180  
 Db 121 AGGTGGAGGACAAAGGTGCTGAGGATTTGGGAGGCGAACCTATGGCAAGTTGGTGAAGAGGC 180  
 QY 181 ACGCAATGAGAGATCTATTCAGACTTACACTGATGCGCGCAACAAATTCACCTTTAGA 240  
 Db 181 ACGCAATGAGAGATCTATTCAGACTTACACTGATGCGCGCAACAAATTCACCTTTAGA 240  
 QY 241 TTTTGATCTGTCACTCCCTACTTATTCCTGGTGGGCAACTTCCCAATAGGCTCATGTT 300  
 Db 241 TTTTGATCTGTCACTCCCTACTTATTCCTGGTGGGCAACTTCCCAATAGGCTCATGTT 300  
 QY 301 AATCAATGATTAGTGTATTTCAGCAAAATATTCCTGTTGTTGTTGACATTTATATATGTG 360  
 Db 301 AATCAATGATTAGTGTATTTCAGCAAAATATTCCTGTTGTTGTTGACATTTATATATGTG 360  
 QY 361 GGGTGAGACGGATTAATAATATCATCCATGAGAGCTTTATCTCTCTGTTGTTGTTGTTG 420  
 Db 361 GGGTGAGACGGATTAATAATATCATCCATGAGAGCTTTATCTCTCTGTTGTTGTTGTTG 420  
 QY 421 TTTTCAGATCATTTCTCAGTGTTCACAGAAATTTCTCAGTGTTCAGTGTTCAGTGTTCAG 480  
 Db 421 TTTTCAGATCATTTCTCAGTGTTCACAGAAATTTCTCAGTGTTCAGTGTTCAGTGTTCAG 480  
 QY 481 AAGTGAGGTTCTTTAAATTTTCATTTATGCTTCTCTTTCTTTCTAGACTAGCAACTGCATGA 540  
 Db 481 AAGTGAGGTTCTTTAAATTTTCATTTATGCTTCTCTTTCTTTCTAGACTAGCAACTGCATGA 540  
 QY 541 CTTTTCATCTTTGGGTTTCACAAATTTGACTCACAAGAAAACAAATTCACCTTTGGGTTTACA 600  
 Db 541 CTTTTCATCTTTGGGTTTCACAAATTTGACTCACAAGAAAACAAATTCACCTTTGGGTTTACA 600  
 QY 601 AATTCCTCTTCAGGATGTACTTTTCACTTGAACGTGATGATAGGAACAGGAAATGGCT 660  
 Db 601 AATTCCTCTTCAGGATGTACTTTTCACTTGAACGTGATGATAGGAACAGGAAATGGCT 660  
 QY 661 CAGTTTTTAAGGAACAATGTACAGATTTCAITTCAGAACTCTTCTGTTGTTGAGTTT 720  
 Db 661 CAGTTTTTAAGGAACAATGTACAGATTTCAITTCAGAACTCTTCTGTTGTTGAGTTT 720  
 QY 721 CAGACTTTTTGTACAAGCTGATGGATCAAAATCTTGTTCCTCAAGCTCTGATTAACAGAA 780  
 Db 721 CAGACTTTTTGTACAAGCTGATGGATCAAAATCTTGTTCCTCAAGCTCTGATTAACAGAA 780  
 QY 781 ACTGGCAACTCCTAATTTGATTAATAAAGAAATAAATACAGTATCAGATATCTCATTTTC 840  
 Db 781 ACTGGCAACTCCTAATTTGATTAATAAAGAAATAAATACAGTATCAGATATCTCATTTTC 840  
 QY 841 TTGGTTGGCAGATCACAAGAAAGGAAACACAAAGGCTTAAGCCCTCTTCTGTTGGGAGTTA 900





CC fertility in a plant e.g. maize. An isolated nucleotide sequence  
CC comprising the SBM200 gene is useful for producing hybrid seeds and for  
CC providing heritable externally controllable male sterility in a plant.  
CC The present sequence represents an essential region of the maize SBM200  
CC promoter sequence

XX Sequence 267 BP; 81 A; 66 C; 54 G; 66 T; 0 U; 0 Other;

Query Match 6.7%; Score 263; DB 6; Length 267;

Best Local Similarity 100.0%; Pred. No. 8.9e-49;

Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 AACTCATTTTCTGTGGTGGCAGATCACAATAAGGACACAAAGGCTAAGCCTCCTACTTG 889

DB 5 ATCTCATTTTCTGTGGTGGCAGATCACAATAAGGACACAAAGGCTAAGCCTCCTACTTG 64

QY 890 TTCCGGAGTTAGTCAAGGACACCATATCAATGAAAGAAATCTTAATTGGGGTCAACAC 949

DB 65 TTCCGGAGTTAGTCAAGGACACCATATGAAAGAAATCTTAATTGGGGTCAACAC 124

QY 950 AAGATTGTCTCTCTCGAGTTGGGGGTCCTTAAGGTTGGTAGTACCAATACCCAATATA 1009

DB 125 AAGATTGTCTCTCTCGAGTTGGGGGTCCTTAAGGTTGGTAGTACCAATACCCAATATA 184

QY 1010 TCACCTAACAAACCCATCCATGCTACATACATACATACATACATCCATCATTGTAGACTGG 1069

DB 185 TCACCTAACAAACCCATCCATGCTACATACATACATACATACATCCATCATTGTAGACTGG 244

QY 1070 ACCCTTCATCAAGAGCACCATTGG 1092

DB 245 ACCCTTCATCAAGAGCACCATTGG 267

## RESULT 6

ADA70596

ID ADA70596 standard; DNA; 1527 BP.

AC ADA70596;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 3919.

XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW Gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.

XX Claim 6; SEQ ID NO 3919; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes

CC involved in plant resistance or response to pathogenic infection. M1

CC comprises identifying a gene whose expression is significantly altered in

CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

XX Sequence 1527 BP; 280 A; 473 C; 494 G; 280 T; 0 U; 0 Other;

Query Match 3.5%; Score 137.8; DB 7; Length 1527;

Best Local Similarity 70.5%; Pred. No. 1.7e-20;

Matches 184; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 2692 GCGCAGGACCCCAAGGGATCCTGGAGGACGACGTCGTGCGGACGGACGAGGTGAGG 2751

DB 1135 GCATGGACTCGCACAGCTGCAAGGAGGACGACTTCTTCCGCGGACGGCACGTTCTGGGG 1194

QY 2752 GCCGCGGGATGGTCACTGACGTGCGCTACTCGATGGGGCGGATGAGTACAACTGGGGC 2811

DB 1195 AAAGGTTGGTTGTGTCTCTACAGCGCTACGCGATGGCGGATCGAGGACATCTGGGGC 1254

QY 2812 CCCGACGCGCGGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGCGCGTTCCGCAAC 2871

DB 1255 GGGACTGTGAGGAGTTTCAAGCCAGAGCGGTGCTAGACGAGCGCGCGTTCCGCGCG 1314

QY 2872 GCGTCGCGCTTCAAGTTCAAGCGGTTCCAGGCGGCGGCGGAGATCTGCTGGGCAAGGAC 2931

DB 1315 GAGAGCCCGTTCAAGTACCGGTTTCCACGCGGCGGCGGAGATGTGCTCGGCAAGGAG 1374

QY 2932 TCGGCGTACCTGCACATGAAG 2952

DB 1375 ATGGCCTACATACATGAAG 1395

## RESULT 7

ADA71342

ID ADA71342 standard; DNA; 1536 BP.

AC ADA71342;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 4665.

XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW Gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.

XX Claim 6; SEQ ID NO 4665; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes



CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

XX  
SQ Sequence 1536 BP; 301 A; 478 C; 479 G; 278 T; 0 U; 0 Other;  
Query Match 3.3%; Score 127; DB 7; Length 1536;  
Best Local Similarity 68.6%; Pred. NO. 4.3e-18;  
Matches 175; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
Qy 2698 GACCCAGGGGATCTTGGAGGACGCTGCTGCGGACGGGACGAGGTGAGGCGCGC 2757  
Db 1147 GACACGCACAGTGCAGAGGAGGACGCTTCTGCGGACGGCAGCTTCGTGGGAAAGG 1206  
Qy 2758 GGGATGGTACGCTACGTCGCCCTACTCGATGGGGGGATGGAGTACAACTGGGGGCCCGAC 2817  
Db 1207 TGGCTAGTGTATACAGCGCGTATGCCATGGGCGTATGGAGGCATCTGGGCGCGGAC 1266  
Qy 2818 GCGGCGAGCTTCCGGCCCGGACGGTGGATCAACGAGGATGGCGCTTCCGCAACGCGTCG 2877  
Db 1267 TCGGAGGAGTACAGGCCGGAACGGTGGCTGGACGAGGCGGCGGCTTCCGCGCGAGAGC 1326  
Qy 2878 CGGTTCAGTTACGGCGCTTCCAGCGGGCGGAGGATCTGCTGGGCAAGGACTCGGCG 2937  
Db 1327 ACGTTCAAGTACCCGGTGTTCAGCGCGGCGCTTGGATCTGATCGGCAAGGATGGCC 1386  
Qy 2938 TACCTGCAGATGAAG 2952  
Db 1387 TACATACAAATGAAG 1401

RESULT 8  
ACL21310/C  
ID ACL21310 standard; DNA; 657 BP.  
XX  
AC ACL21310;  
XX  
XX  
DT 27-OCT-2003 (revised)  
DT 17-OCT-2003 (first entry)  
XX  
DE DNA clone originating in barley containing SNP encoding sequence #11301.  
XX  
KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
KW gene; ss.  
OS Hordeum vulgare; var. (cul.Akashinriki).  
XX  
XX WO2003057877-A1.  
XX  
XX 17-JUL-2003.  
XX  
XX 16-DEC-2002; 2002WO-IB005403.  
XX  
XX 20-DEC-2001; 2001JP-00387059.  
XX  
XX 20-DEC-2001; 2001JP-00387131.  
XX  
XX 20-DEC-2001; 2001JP-00403299.  
XX  
XX 20-DEC-2001; 2001JP-00403300.  
XX  
XX 27-SEP-2002; 2002JP-00327515.  
XX  
XX (UYN1-) UNIV JAPAN OKAYAMA.  
XX  
XX Sato K, Takeda K, Kohara Y;  
XX WPI; 2003-587127/55.  
XX  
XX

Single nucleotide polymorphism sites in barley varieties and DNA  
sequences containing them for analysis and identification of barley

PT varieties and production of barley transformants with desired  
PT characteristics.  
XX  
XX Disclosure; SEQ ID XX; 284pp; Japanese.  
XX  
CC The present invention relates to oligonucleotide clones originating in  
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
CC varieties, identification of particular varieties and genotype-phenotype  
CC analysis, isolation of specific genes and creation of new varieties by  
CC transformation of barley varieties with them and production of new barley  
CC varieties with desired properties. The present sequence represents an  
CC oligonucleotide clone DNA sequence featured in the specification. The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 657 BP; 140 A; 227 C; 159 G; 131 T; 0 U; 0 Other;

Query Match 3.1%; Score 120.8; DB 8; Length 657;  
Best Local Similarity 60.7%; Pred. No. 7.7e-17;  
Matches 216; Conservative 0; Mismatches 137; Indels 3; Gaps 1;  
Qy 2713 CTGAGAGACACGCTGCTGCGGACGGACGAGGTGAGGCGCGGATGGTGACGTAC 2772  
Db 560 CAGGCCCGACGCTTCTTCCGACGGCACCGCGGTGGGACCGGTTGGTTCGTGGCGTAC 501  
Qy 2773 GTGCGCTACTCGATGGGGCGGATGGAGTCAAACTGGGGCCCCGACGCGGCGAGCTTCGG 2832  
Db 500 AACTCTGTATGCAATGGGCGGATGGAGTCTGTGTGGGCGGAGACGCGAGGCGTACCGG 441  
Qy 2833 CCGGAGCGGTGGATCAAC---GAGGATGGCGCGTTCGCAACGCGTCCCGTTCAGTTC 2889  
Db 440 CCGGAGCGGTGGTGGACCCGCGGCGGAGACGTTCCGCGCGGAGAGCCGTTCCGCTAC 381  
Qy 2890 ACGCGCTTCAGGCGGGCGGAGATCTGCTGGCAAGGACTCGGCGTACCTGCAGATG 2949  
Db 380 ATGGCGTTTCAACGCGGGCGCAAGATTGCTTCGAAAGGAGATGGCGTATATCCAGATG 321  
Qy 2950 AAGATGGCGTGGCCATCTTTCGCGCTTCTACAGCTTCCGCGTCTGAGGGGCGACCCG 3009  
Db 320 AAGTCTATCGTGGCGTGGTGGTGGAGAGTTTGAATTGGCGGTGGACGCGCGTACCGG 261  
Qy 3010 GTGCAGTACCGCATGATGACCATCTCTCCATGGCGCACGCGCTCAAGGTCGCGGT 3065  
Db 260 CCGCGGCGAGTGGCATCACTCACTTACGAATGGCGGACGGGCTCCGCGTCCGCGGT 205

RESULT 9  
AA26878  
ID AA26878 standard; cDNA; 1626 BP.  
XX  
XX  
AC AA26878;  
XX  
XX  
DT 23-JUN-1999 (first entry)  
XX  
XX  
DE Codon modified cDNA encoding a cytochrome P450.  
XX  
XX Cytochrome P450; protein expression; yeast; mutant; ss.  
XX  
XX Synthetic.  
XX OS Triticum aestivum.  
XX  
XX Key Location/Qualifiers  
XX CDS 1..1626  
XX /\*tag= a  
XX  
XX FR2768748-A1.  
XX  
XX  
XX 26-MAR-1999.  
XX  
XX 24-SEP-1997; 97FR-00012094.  
XX  
XX



ADA70663  
 ID ADA70663 standard; DNA; 1563 BP.  
 XX AC  
 XX ADA70663;  
 XX XX  
 DT 20-NOV-2003 (first entry)  
 XX XX  
 DE Rice gene, SEQ ID 3986.  
 XX XX  
 KW plant; bacterial infection; fungal infection; viral infection; rice;  
 KW gene; ds.  
 XX XX  
 OS Oryza sativa.  
 XX XX  
 FN WO2003000898-A1.  
 XX XX  
 PD 03-JAN-2003.  
 XX XX  
 EF 22-JUN-2001; 2001WO-IB001105.  
 XX XX  
 PR 22-JUN-2001; 2001WO-IB001105.  
 XX XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX XX  
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 XX XX  
 DR WPI; 2003-175290/17.  
 XX XX  
 PT Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX XX  
 PS Claim 6; SEQ ID NO 3986; 899pp; English.  
 XX XX  
 CC The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX XX  
 SQ Sequence 1563 BP; 221 A; 572 C; 530 G; 240 T; 0 U; 0 Other;  
 Query Match 2.9%; Score 114; DB 7; Length 1563;  
 Best Local Similarity 52.6%; Pred. No. 3.5e-15;  
 Matches 320; Conservative 0; Mismatches 280; Indels 8; Gaps 3;  
 QY 2406 CTGTGCGCTTCGAGCGGAGCGCGCGCGGAGGAGGCGTGTGCTGCTGCGCGC 2465  
 DB 850 CTGTGCGGTTTCATGGCTTCCACCGGCTGCGAGCGCGCGCGGAGCAGCAAGTTCCTC 909  
 QY 2466 GCGCTGACGCGCAGACAGAGCGGTTCGCGCGCGCGGTGCGCGGCTTCCTC 2525  
 DB 910 CGCGACATCGTGTGAGTTCCTCTCGCGCGCGGAGACAGGTGTCCACCGGTCA 969  
 QY 2526 ACCTAGCAGCTCGGAGCTGTGTCTACCTCCAGCGTGTGTCACCGAGCGTCCG 2585  
 DB 970 AGCTGTTCATGCTCTCTGTCCAGAACCCCGAGTGTGCGCGGCTATGCGCGGAGCG 1029  
 QY 2586 CTGTACCCCGCGCTTCCTCAGGTGAGCGCGCGCGCAGCAGCGCTCGGTCCAGAG--- 2641  
 DB 1030 GAGCGCGGAGCGGCGGAGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCTC 1089  
 QY 2642 CACAGCATGCAATGAGTGTGAGCTGATGATGCAATGCAATGCAATGCGGCGC- 2700  
 DB 1090 CACTACACCCAGCGCGTGTGTCACGAGAACATGCGGCTGTTCGCCGCGGTGCGATTCGAC 1149

2701 CCCAAGGGGATCTTGAGAGGACGACGTCTGCCCGACCGGACGAGGTGAGGGCCGCGCGG 2760  
 DB 1150 TCCAAAGTTTTCGCGCGCGCGCGACGTCTGCCCGACCGCACCTACGTGCGGCGGACGCG 1209  
 QY 2761 ATGCTGACGTACGTGCCCTACTCGATGGGCGGATGAGTACAACTGGGCGCCGACGCG 2820  
 DB 1210 CGGTATGATACACCCCTACGCCATGGGCGGATGCCGACATCTGGGCGCGGATAC 1269  
 QY 2821 GCGAGCTTCGCGCGGAGCGGTGATCAACGAGGATGCG---CGTTCCGCAAGCGTCG 2877  
 DB 1270 GCGCGTTCCGCGCGCGCGGTGCTCACCGCGCGCGCGTCTGTTCTGTTCCGCGCAAC 1329  
 QY 2878 CCGTTCAAGTTCACGCGGTTCCAGGCGGCGCGGAGATCTCCCTGGGCAAGGACTCGCG 2937  
 DB 1330 CCGTACAAATACCCGGTGTTCAGGCGCGGCAACGCGTGTGCTCGCAAGGAGCTCGCC 1389  
 QY 2938 TACCTGCAGATGAAGATGGGCTGGCCATCTCTTGGGCTTCTACAGTTCGCGCTGCTG 2997  
 DB 1390 GTACCCGAGATGAAGCGGCGGCGGCGGCTGCTGAGGGGTTTCGAGCTCGAGTCTGTC 1449  
 QY 2998 GAGGGGCA 3005  
 DB 1450 GCGGAGAA 1457

RESULT 12  
 ACL21306/c  
 ID ACL21306 standard; DNA; 621 BP.  
 XX AC  
 AC ACL21306;  
 XX XX  
 DT 27-OCT-2003 (revised)  
 DT 17-OCT-2003 (first entry)  
 XX XX  
 DE DNA clone originating in barley containing SNP encoding sequence #11297.  
 XX XX  
 KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
 KW Gene; ss.  
 XX XX  
 OS Hordeum vulgare; var. (cul.Akashinriki).  
 XX XX  
 PN WO2003057877-A1.  
 XX XX  
 PD 17-JUL-2003.  
 XX XX  
 PF 16-DEC-2002; 2002WO-IB005403.  
 XX XX  
 PR 20-DEC-2001; 2001JP-00387059.  
 PR 20-DEC-2001; 2001JP-00387131.  
 PR 20-DEC-2001; 2001JP-00403299.  
 PR 20-DEC-2001; 2001JP-00403300.  
 PR 27-SEP-2002; 2002JP-00327515.  
 XX XX  
 PA (UYN1-) UNIV JAPAN OKAYAMA.  
 XX XX  
 PI Sato K, Takeda X, Kohara Y;  
 XX XX  
 DR WPI; 2003-587127/55.  
 PT Single nucleotide polymorphism sites in barley varieties and DNA  
 PT sequences containing them for analysis and identification of barley  
 PT varieties and production of barley transformants with desired  
 PT characteristics.  
 XX XX  
 PS Disclosure; SEQ ID XX; 284pp; Japanese.  
 XX XX  
 CC The present invention relates to oligonucleotide clones originating in  
 CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
 CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
 CC varieties, identification of particular varieties and genotype-phenotype  
 CC analysis, isolation of specific genes and creation of new varieties by  
 CC transformation of barley varieties with them and production of new barley  
 CC varieties with desired properties. The present sequence represents an

XX	Disclosure; SEQ ID NO 1380; 299pp; English.
XX	
CC	The invention relates to a novel isolated polynucleotide comprising a
CC	plant nucleotide sequence having an open reading frame that encodes a
CC	polypeptide associated with disease resistance or its fragment having
CC	substantially the same activity as the full-length polypeptide. The
CC	polynucleotide of the invention is useful for conferring resistance or
CC	tolerance to a plant pathogen. The present sequence represents a gene
CC	conferring disease resistance used in the invention.
XX	
SQ	Sequence 528 BP; 75 A; 191 C; 183 G; 79 T; 0 U; 0 Other;
	Query Match                    2.7%;   Score 105.4;   DB 8;   Length 528;
	Best Local Similarity       61.2%;   Pred. No. 2e-13;
	Matches 188; Conservative     0; Mismatches 116; Indels     3; Gaps     1;
QY	2698   GACCCCAAGGGGATCTGGAGGACGACGTGCTGCCGGACCGGACGAGGTGAGGGCCGGC   2757
Db	124   GACTCCAAGTTCTTCGTGGGGCCAGACGTGCTCCCAGACGGCACCTACGTGCCGGGGCGGC   183
QY	2758   GGGATCGTGACGTACGTGCCCTACTCGATGGGCGGATGGAGTACAACCTGGGGCCCCGCAC   2817
Db	184   ACGGGCTGACCTACCACCCCTACGCATGGGCGCGCATGCCGGCGATCTGGGGCGCGAC   243
QY	2818   GCGGCGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGGCG--CGTTCCGAAACGG   2874
Db	244   CACGGGCGCTTCGCGCGCGGCGCGCTGGCTCACCGGCGCGCGGGCTGTTCTGCTCCCGAG   303
QY	2875   TCGCCGCTCAAGTTACGGGTTCCAGGGGGGGCCGAGGATCTGCTGGGCAAGGACTCG   2934
Db	304   AGCCTGTACAAGTACCGGTGTTCCAGCGGGGCTTCGCGTGTGCCTCGGCACAGAGCTC   363
QY	2935   CGGTACTCGAGATGAAGATGGCGCTGGGCATCCTCTTGGCGCTTCTACAGCTTCCGGCTG   2994

Qy	2995	CTGGAGG	3001
Db	424	GTCGGG	430

RESULT 14

ACL21305/c	
ID	ACL21305 standard; DNA; 399 BP.
XX	
XX	ACL21305;
XX	
XX	27-OCT-2003 (revised)
DT	17-OCT-2003 (first entry)
XX	
XX	DNA clone originating in barley containing SNP encoding sequence #11296.
DE	Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KW	gene; ss.
KW	
XX	
OS	Hordeum vulgare; var. (cul.Akashinriki).
XX	
FN	WO2003057877-A1.
XX	
PD	17-JUL-2003.
XX	
PF	16-DEC-2002; 2002WO-IB005403.
XX	
FR	20-DEC-2001; 2001JP-00387059.
PR	20-DEC-2001; 2001JP-00387131.
PR	20-DEC-2001; 2001JP-00403299.
PR	20-DEC-2001; 2001JP-00403300.
PR	27-SEP-2002; 2002JP-00327515.
XX	
PA	(UYNI-) UNIV JAPAN OKAYAMA.
XX	
PI	Sato K., Takeda K., Kohara Y.

```
XX DR WPI; 2003-587127/55.
XX PF
XX PR
XX PR Single nucleotide polymorphism sites in barley varieties and DNA
XX PT sequences containing them for analysis and identification of barley
XX PT varieties and production of barley transformants with desired
XX PT characteristics.
XX PS
XX PS Disclosure; SEQ ID XX; 284pp; Japanese.
XX PA
XX CC The present invention relates to oligonucleotide clones originating in
XX CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX CC varieties, identification of particular varieties and genotype-phenotype
XX CC analysis, isolation of specific genes and creation of new varieties by
XX CC transformation of barley varieties with them and production of new barley
XX CC varieties with desired properties. The present sequence represents an
XX CC oligonucleotide clone DNA sequence featured in the specification. The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
XX CC standardise OS field)
XX CC
XX CC Sequence 399 BP; 74 A; 162 C; 86 G; 77 T; 0 U; 0 Other;
XX CC
XX CC Query Match 2.7%; Score 104; DB 8; Length 399;
XX CC Best Local Similarity 59.5%; Pred. No. 3.8e-13;
XX CC Matches 195; Conservative 0; Mismatches 130; Indels 3; Gaps 1;
XX CC
XX CC 2741 CGAAGTGAGGCGCGCGGATGGTACGTACGTCCCTACTCGATGGGCGGATGGAGT 2800
XX CC Db
XX CC 398 CGCGGTGGGACGGGTGGTTCGTGGCGTACACTCGTATGCATGGGCGGATGGAGT 339
XX CC Qy
XX CC 2801 ACAACTGGGCGCCGACCGCGGAGCTTCGGCGCGGAGCGGTGGATCAAC---GAGGATG 2857
XX CC Db
XX CC 338 CTGTGTGGGGGAGACGCGAGGCGGTACCGCGCGGAGCGGTGGTGGACCCGCGGAGG 279
XX CC Qy
XX CC 2858 GCGGTTCGCGAACGCGTCGCGTTCAGTTCACGCGGTTCCAGCGGCGCGGAGTCT 2917
XX CC Db
XX CC 278 GGACGTTCCGCGCGGAGAGCGCGTTCGCGTACATGGCGTTTCACCGGGGCGCAAGATT 219
XX CC Qy
XX CC 2918 GCCTGGCAAGACTCGCGTACCTGCAGATGAAGATGGCGTGGCCATCTCTTGGCGCT 2977
XX CC Db
XX CC 218 GCCTCGAAGAGAGATGGCGTATATCCAGATGAAGTCTATCGTGGCGTGGCTGGAGG 159
XX CC Qy
XX CC 2978 TCTACAGCTTCGCGTCTGGAGGGGACCCGCGTGCAGTACCGCATGATGACCATCTCT 3037
XX CC Db
XX CC 158 AGTTTGAATTGGCGGTGGACCGCGGTACCGCGCGGAGGTGGCATCTCTGACGTTAC 99
XX CC Qy
XX CC 3038 CCATGGCGCAGCGCTCAAGGTCGCGGT 3065
XX CC Db
XX CC 98 GGATGGCGACGGGCTCCCGGTGGCGGT 71
XX CC
XX CC RESULT 15
XX CC ACL21309/c
XX CC ID ACL21309 standard; DNA; 460 BP.
XX CC AC
XX CC AC ACL21309;
XX CC AC
XX CC DT 27-OCT-2003 (revised)
XX CC DT 17-OCT-2003 (first entry)
XX CC DE
XX CC DE DNA clone originating in barley containing SNP encoding sequence #11300.
XX CC KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
XX CC KW gene; ss.
XX CC OS Hordeum vulgare; var. (cul.Akashinriki).
XX CC PN WO2003057877-A1.
XX CC XX
XX CC PD 17-JUL-2003.
```

```
XX PF 16-DEC-2002; 2002WO-IB005403.
XX XX
XX PR 20-DEC-2001; 2001JP-00387059.
XX PR 20-DEC-2001; 2001JP-00387131.
XX PR 20-DEC-2001; 2001JP-00403299.
XX PR 20-DEC-2001; 2001JP-00403300.
XX PR 27-SEP-2002; 2002JP-00327515.
XX PA
XX PA (UYNI-) UNIV JAPAN OKAYAMA.
XX PI Sato K, Takeda K, Kohara Y;
XX XX
XX XX WPI; 2003-587127/55.
XX XX
XX PT Single nucleotide polymorphism sites in barley varieties and DNA
XX PT sequences containing them for analysis and identification of barley
XX PT varieties and production of barley transformants with desired
XX PT characteristics.
XX PT
XX PS Disclosure; SEQ ID XX; 284pp; Japanese.
XX XX
XX CC The present invention relates to oligonucleotide clones originating in
XX CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX CC varieties, identification of particular varieties and genotype-phenotype
XX CC analysis, isolation of specific genes and creation of new varieties by
XX CC transformation of barley varieties with them and production of new barley
XX CC varieties with desired properties. The present sequence represents an
XX CC oligonucleotide clone DNA sequence featured in the specification. The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
XX CC standardise OS field)
XX CC
XX CC Sequence 460 BP; 95 A; 174 C; 98 G; 91 T; 0 U; 2 Other;
XX CC
XX CC Query Match 2.7%; Score 104; DB 8; Length 460;
XX CC Best Local Similarity 59.5%; Pred. No. 3.8e-13;
XX CC Matches 195; Conservative 0; Mismatches 130; Indels 3; Gaps 1;
XX CC
XX CC 2741 CGAAGTGAGGCGCGCGGATGGTACGTACGTCCCTACTCGATGGGCGGATGGAGT 2800
XX CC Db
XX CC 459 CGCGGTGGGACGGGTGGTTCGTGGCGTACACTCGTATGCATGGGCGGATGGAGT 400
XX CC Qy
XX CC 2801 ACAACTGGGCGCCGACCGCGGAGCTTCGGCGCGGAGCGGTGGATCAAC---GAGGATG 2857
XX CC Db
XX CC 399 CTGTGTGGGGGAGACGCGAGGCGGTACCGCGCGGAGCGGTGGCTGGACCCGCGGAGG 340
XX CC Qy
XX CC 2858 GCGGTTCGCGAACGCGTCGCGTTCAGTTCACGCGGTTCCAGCGGCGCGGAGGATCT 2917
XX CC Db
XX CC 339 GGACGTTCCGCGCGGAGAGCGCGTTCGCGTACATGGCGTTTCACCGGGGCGCAAGATT 280
XX CC Qy
XX CC 2918 GCCTGGCAAGACTCGCGTACCTGCAGATGAAGATGGCGTGGCCATCTCTTGGCGCT 2977
XX CC Db
XX CC 279 GCCTCGGAAGAGAGATGGCGTATATCCAGATGAAGTCTATCGTGGCGTGGCTGGAGG 220
XX CC Qy
XX CC 2978 TCTACAGCTTCGCGCTCTGAGGGGACCCGCGTGCAGTACCGCATGATGACCATCTCT 3037
XX CC Db
XX CC 219 AGTTTGAATTGGCGGTGGACCGCGGTACCGCGCGGAGGTGGCATCTCTGACGTTAC 160
XX CC Qy
XX CC 3038 CCATGGCGCAGCGCTCAAGGTCGCGGT 3065
XX CC Db
XX CC 159 GGATGGCGGACGGGCTCCCGGTGGCGGT 132
XX CC
XX CC Search completed: March 7, 2004, 12:36:24
XX CC Job time : 968 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2004, 10:17:57 ; Search time 9777 Seconds  
(without alignments)  
17276.036 Million cell updates/sec

Title: US-10-021-657-7  
Perfect score: 3897  
Sequence: 1 gaattccaagcaggccctt.....gctgagtaaacgacgaattc 3897

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3897	100.0	3897	6	AX453578 Sequence
2	1382.2	35.5	1891	8	AF366297 Zea mays
3	1378.4	35.4	1906	6	AX453572 Sequence
4	1114.8	28.6	160738	8	AC073556 Oryza sat
5	1092	28.0	1092	6	AX453576 Sequence
6	1051	28.0	1091	8	AF366296 Zea mays
7	375	9.6	107214	8	AC073178 Arabidops
8	268.4	6.9	494	6	AX453574 Sequence
9	263	6.7	267	6	AX453577 Sequence
10	189.2	4.9	116932	8	AC074232 Oryza sat
11	189.2	4.9	300029	8	AE017114 Oryza sat
12	186	4.8	1695	8	AK118803 Arabidops
13	174.4	4.5	140681	8	AC078894 Oryza sat
14	174.4	4.5	300029	8	AE017113 Oryza sat
15	162.4	4.2	146519	8	CNS08C9V Oryza sat
16	162.4	4.2	174521	8	CNS08C8D Oryza sat
17	156.8	4.0	1881	8	AK061650 Oryza sat
18	155.4	4.0	154865	8	OSJN00071 Oryza sat
19	151	3.9	1935	8	AK107532 Oryza sat
20	151	3.9	134821	8	AC137697 Genomic s
21	151	3.9	156647	8	AC137930 Oryza sat
22	144.6	3.7	1746	8	AK109775 Oryza sat
23	144.6	3.7	115787	2	AP004092 Oryza sat
24	143.8	3.7	1714	8	AK119612 Oryza sat
25	143.8	3.7	1780	8	AK103182 Oryza sat
26	143.8	3.7	1904	8	AK101618 Oryza sat
27	142.2	3.6	1712	8	AK120901 Oryza sat
28	142	3.6	1757	8	AK070931 Oryza sat
29	142	3.6	165497	8	AP003232 Oryza sat
30	141.4	3.6	1120	8	AK061292 Oryza sat
31	140.2	3.6	1371	8	AK105884 Oryza sat
32	140.2	3.6	1658	8	AK119195 Oryza sat
33	140.2	3.6	1916	8	AK104799 Oryza sat
34	140.2	3.6	1933	8	AK069358 Oryza sat
35	138.6	3.6	1549	8	AK067844 Oryza sat
36	138.6	3.6	1934	8	AK099925 Oryza sat
37	137.8	3.5	1527	6	AX654049 Sequence
38	137.8	3.5	1532	8	AK064306 Oryza sat
39	137.6	3.5	104307	8	CNS08CDU Oryza sat
40	137.6	3.5	132026	8	CNS08CDS Oryza sat
41	136.6	3.5	1564	8	AK102714 Oryza sat
42	136	3.5	1928	8	AK103307 Oryza sat
43	136	3.5	141025	2	AC123528 Oryza sat
44	131	3.4	160350	2	AC123527 Oryza sat
45	130.2	3.3	1448	8	AK109052 Oryza sat

ALIGNMENTS

RESULT 1  
AX453578  
LOCUS AX453578 3897 bp DNA linear PAT 02-SEP-2002  
DEFINITION Sequence 7 from Patent WO0226789.  
ACCESSION AX453578  
VERSION AX453578.1 GI:21712820  
KEYWORDS  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1  
AUTHORS Albertsen, M.C., Fox, T., Huffman, G. and Trimmell, M.  
TITLE Nucleotide sequences mediating male fertility and method of using





QY 1981 AGAGGCCCTCTAGGCGAGAGCATCAAGCTCGTGGACGAGTTCACCTACAGGCTGATCCG 2040  
DB 1981 AGAGGCCCTCTAGGCGAGAGCATCAAGCTCGTGGACGAGTTCACCTACAGGCTGATCCG 2040  
QY 2041 CCGGAGGAAGCCGAGATCTGTCGAGGCCCGGGCCAGCGGCAAAACAGGAGAAAGGTACGTGC 2100  
DB 2041 CCGGAGGAAGCCGAGATCTGTCGAGGCCCGGGCCAGCGGCAAAACAGGAGAAAGGTACGTGC 2100  
QY 2101 ACATGACTGTTTCGATTCTTCAGTTCATCTGCTTGGCCGGGATGACCTGATCTGATGTTG 2160  
DB 2101 ACATGACTGTTTCGATTCTTCAGTTCATCTGCTTGGCCGGGATGACCTGATCTGATGTTG 2160  
QY 2161 ATTATATATCCGTGTGACTTGTGAGGACAAATTAATTAATGAGGACGATGAGCAGCATCC 2220  
DB 2161 ATTATATATCCGTGTGACTTGTGAGGACAAATTAATTAATGAGGACGATGAGCAGCATCC 2220  
QY 2221 TGTACAGGTTTCATCGAGCTAGCGAGGCGCGGACGACGCGGCGGCTTCGGGGACGACA 2280  
DB 2221 TGTACAGGTTTCATCGAGCTAGCGAGGCGCGGACGACGCGGCGGCTTCGGGGACGACA 2280  
QY 2281 AGAGCTTCGGGACGCTGCTGCTCAATCTGCTGATCGCCGGGGGGGACGACGCGGCGACA 2340  
DB 2281 AGAGCTTCGGGACGCTGCTGCTCAATCTGCTGATCGCCGGGGGGGACGACGCGGCGACA 2340  
QY 2341 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400  
DB 2341 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400  
QY 2401 GCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2460  
DB 2401 GCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2460  
QY 2461 GCGGCGGCGCTGACGCGGACGACGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520  
DB 2461 GCGGCGGCGCTGACGCGGACGACGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520  
QY 2521 TCTCAGCTTACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580  
DB 2521 TCTCAGCTTACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580  
QY 2581 TCGGCTGTACCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640  
DB 2581 TCGGCTGTACCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640  
QY 2641 GCACAGCATGCACTGAGTGGACCTGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 2700  
DB 2641 GCACAGCATGCACTGAGTGGACCTGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 2700  
QY 2701 CCCAAGGGGATCCTGAGGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760  
DB 2701 CCCAAGGGGATCCTGAGGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760  
QY 2761 ATGCTGACGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820  
DB 2761 ATGCTGACGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820  
QY 2821 GCGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGCGCGGCTGCGCGGCTGCGCGGCTGCGCGG 2880  
DB 2821 GCGAGCTTCGCGCGGAGCGGTGGATCAACGAGGATGCGCGGCTGCGCGGCTGCGCGGCTGCGCGG 2880  
QY 2881 TTCAAGTTTCAAGGCTTCCAGGCGGGGCGGAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940  
DB 2881 TTCAAGTTTCAAGGCTTCCAGGCGGGGCGGAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940  
QY 2941 CTGCAGATGAAGTGGCGCTGGCCATCTCTTGGCTTCTACAGCTTCCGCTGCTGCTGCTGCTGCTG 3000  
DB 2941 CTGCAGATGAAGTGGCGCTGGCCATCTCTTGGCTTCTACAGCTTCCGCTGCTGCTGCTGCTGCTG 3000  
QY 3001 GGGCACCCGCTGAGTACCGCATGATGACCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060  
DB 3001 GGGCACCCGCTGAGTACCGCATGATGACCATCTCTCTCCATGGCGCACCGGCTCAAGGTC 3060  
QY 3061 CGCGTCTTAGGGCGCTGATGCTAGCGGATTTGGGATATCATCCCGCTTAATCTCTTA 3120

DB 3061 CGCGTCTTAGGGCGCTGATGCTAGTTCATGGGATTTGGGATATCATCCGCTTAATCTCTTA 3120  
QY 3121 AAAATTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3180  
DB 3121 AAAATTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3180  
QY 3181 CCTTAAAAATCTCCGCTCGGCTTCGGAACACCAACCACTACTAGTGTGTTTGTACTCTACTC 3240  
DB 3181 CCTTAAAAATCTCCGCTCGGCTTCGGAACACCAACCACTACTAGTGTGTTTGTACTCTACTC 3240  
QY 3241 CTCAGTGGAGTGTAGTGACAGCATACAAGTTCATATATATATATATATATATATATATATATAT 3300  
DB 3241 CTCAGTGGAGTGTAGTGACAGCATACAAGTTCATATATATATATATATATATATATATATATAT 3300  
QY 3301 CGGATGCTTCGCGGACCTTTTGGAGCCTTCTGAGAGCCTTCTGAGAGGCTGTGAAAAAAGGCTTC 3360  
DB 3301 CGGATGCTTCGCGGACCTTTTGGAGCCTTCTGAGAGCCTTCTGAGAGGCTGTGAAAAAAGGCTTC 3360  
QY 3361 TTCTGCGGCGAAGTTTTGGGTTTCAGAGTCTTGGGCTCTTTGCGAGCAAAAAAGGTTTGG 3420  
DB 3361 TTCTGCGGCGAAGTTTTGGGTTTCAGAGTCTTGGGCTCTTTGCGAGCAAAAAAGGTTTGG 3420  
QY 3421 AAGGATCTGAACCTGAAACCGAAATGCTTCGGAATATGCTGCAATCGGCGGCGGCGG 3480  
DB 3421 AAGGATCTGAACCTGAAACCGAAATGCTTCGGAATATGCTGCAATCGGCGGCGGCGG 3480  
QY 3481 GTCACTCGGAGTACGACAGCCCAAGAGCCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
DB 3481 GTCACTCGGAGTACGACAGCCCAAGAGCCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
QY 3541 AGACACTCTCGGACCCCTCGCGCTCGCGAGCTCATCTTCGCTCTCTCTGCTGCTGCTGCTGCTG 3600  
DB 3541 AGACACTCTCGGACCCCTCGCGCTCGCGAGCTCATCTTCGCTCTCTCTGCTGCTGCTGCTGCTG 3600  
QY 3601 TGGCGGACCCGCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3660  
DB 3601 TGGCGGACCCGCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3660  
QY 3661 GTACAAACCCCTCATCTCGCGGCGCGCGAGTACAGCGCCCGGAGCGCGCTATACAT 3720  
DB 3661 GTACAAACCCCTCATCTCGCGGCGCGCGAGTACAGCGCCCGGAGCGCGCTATACAT 3720  
QY 3721 CTATAATCATGATTTGATCTTTTAAACCGGCTTAAACCGGCTTAAACCGGCTTAAACCGGCTTAA 3780  
DB 3721 CTATAATCATGATTTGATCTTTTAAACCGGCTTAAACCGGCTTAAACCGGCTTAAACCGGCTTAA 3780  
QY 3781 GTAAACAGCTTCAAAATTGACACAAATTTAAACAGGACAAACCGTAGCTAAACAGATAG 3840  
DB 3781 GTAAACAGCTTCAAAATTGACACAAATTTAAACAGGACAAACCGTAGCTAAACAGATAG 3840  
QY 3841 AGAATGAGAGACACCCAAAGGTTAGAGATGAAATAGCTGAGTAAACAGCAATTC 3897  
DB 3841 AGAATGAGAGACACCCAAAGGTTAGAGATGAAATAGCTGAGTAAACAGCAATTC 3897

## RESULT 2

AF366297  
LOCUS Zea mays cytochrome P450-like protein (ms\*-8b200) mRNA, complete cds.  
DEFINITION  
AF366297.1 GI:14030556  
ACCESSION  
AF366297  
VERSION  
AF366297.1  
KEYWORDS  
Ze mays  
Ze mays  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 1891)  
AUTHORS Fox, T.W., Trimmell, M.R. and Albertsen, M.C.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2001) Trait and Technology Development, Pioneer



Db 1320 GTACGTGCCCTACTCGATGGGGCGGATGGAGTCAACTGGGGCCCGGACGGCGAGCTT 1379  
Qy 2829 CCGGCGGAGCGGTGGATCAACGAGGATGGCGCGTTCGCCAACGCGTCCCGCTTCAAGTT 2888  
Db 1380 CCGGCGGAGCGGTGGATCAACGAGGATGGCGCGTTCGCCAACGCGTCCCGCTTCAAGTT 1439  
Qy 2889 CACGCGCTTCAGGGCGGCGGAGGATCTGCTGGGCAAGGACTCGGCGTACTTGCAGAT 2948  
Db 1440 CACGCGCTTCAGGGCGGCGGAGGATCTGCTGGGCAAGGACTCGGCGTACTTGCAGAT 1499  
Qy 2949 GAAGATGGCGTGGCATCTCTTCGCTTCTACAGCTTCGCGCTGCTGGAGGGGACCC 3008  
Db 1500 GAAGATGGCGTGGCATCTCTTCGCTTCTACAGCTTCGCGCTGCTGGAGGGGACCC 1559  
Qy 3009 GGTGAGTACCGCATGATGACCATCTCTCCATGGCGCAAGGCTTCAAGGTCGCGTCTC 3068  
Db 1560 GGTGAGTACCGCATGATGACCATCTCTCCATGGCGCAAGGCTTCAAGGTCGCGTCTC 1619  
Qy 3069 TAGGCGCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3117  
Db 1620 TAGGCGCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1679  
Qy 3118 -----TTAAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3164  
Db 1680 AATAAGCTCGTGTACAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1739  
Qy 3165 GGTGCTTGGCTTAAGCCCTTAAAAAATCCGTCGGGCTTTGGAACCAACACATCATAGT 3224  
Db 1740 GGTGCTTGGCTTAAGCCCTTAAAAAATCCGTCGGGCTTTGGAACCAACACATCATAGT 1799  
Qy 3225 GTTTGTACTCTACTCTCAGTGGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 3284  
Db 1800 GTTTGTACTCTACTCTCAGTGGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTATAT 1859  
Qy 3285 TATCCCTTTCTT 3297  
Db 1860 TATCCCTTTCTT 1872

RESULT 3  
AX453572  
LOCUS AX453572 1906 bp DNA linear PAT 02-SBP-2002  
DEFINITION Sequence 1 from Patent WO0226789.  
ACCESSION AX453572  
VERSION AX453572.1 GI:21712815  
KEYWORDS  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1  
Albertsen, M.C., Fox, T., Huffman, G. and Trimmell, M.  
Nucleotide sequences mediating male fertility and method of using  
same  
Patent: WO 0226789-A 1 04-APR-2002;  
PIONEER HI-BRED INTERNATIONAL, INC. (US)  
FEATURES  
Location/Qualifiers  
1. .1906  
/organism="Zea mays"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4577"  
join(<1.1638,1642..1770)  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD38131.1"  
/db\_xref="GI:21712816"  
/db\_xref="REFSEQ:CAD38131"  
/translation="EFQTRAHLTPTPSPFFLAGHKVIALLLVVLVILVQMSL  
RKQKFRSPVIGATVEQLRNHRMDLVGLSRHRTVDMFPTSYTADPVAIVE  
HVLKTNFTNPKGIVTVSRMDVLLDGIENADGELWRKORTASFFASKNLRFSAI  
VEREYSIKLAGIISQASKACKVMDQMLYRMTLDSICKVFGVEIGTILSPDLPENSF  
AQAFDANIIITIRFDPLRIKRFHFVHGSEALLAQSIKLVDSEFTSYVIRRRKAEIVE

CDS

## ORIGIN

Query Match 35.4%; Score 1378.4; DB 6; Length 1906;  
Best Local Similarity 82.1%; Pred. No. 3e-198;  
Matches 1833; Conservative 0; Mismatches 16; Indels 385; Gaps 6;  
Qy 1088 CATGAGGAACTCACATCAGCCGCGGACGCCATCGCCATTTCTCCACTAGCAGGCCC 1147  
Db 9 CACGAGGAACTCACCTCAGCCGCGGACGCCATCGCCATTTCTCCACTAGCAGGCCC 68  
Qy 1148 TCACAAGTACATCGCGTCTCTCTGTTGTTCTCTCATGATCTCTGTTCCAGAGTGAG 1207  
Db 69 TCACAAGTACATCGCGTCTCTCTGTTGTTCTCTCATGATCTCTGTTCCAGAGTGAG 128  
Qy 1208 CTTGAGGAAAGCAAGAGCCGAGATCATGCCAGTCTCGGTGACCAAGTGGAGCAGCT 1267  
Db 129 CTTGAGGAAAGCAAGAGCCGAGATCATGCCAGTCTCGGTGACCAAGTGGAGCAGCT 188  
Qy 1268 GAGGAATACCAACCGGATGACGACTGGCTTGTCTGGGTACCTGTCCAGGCAAGGACAGT 1327  
Db 189 GAGGAATACCAACCGGATGACGACTGGCTTGTCTGGGTACCTGTCCAGGCAAGGACAGT 248  
Qy 1328 GACCGTCAAGATGCGGTTCACTTCTCAACCTACATCGTGAACCGGTGAATGTGAGCA 1387  
Db 249 GACCGTCAAGATGCGGTTCACTTCTCAACCTACATCGTGAACCGGTGAATGTGAGCA 308  
Qy 1388 TGTCTCAAGTAACTTACCAATTTACCCAAATGAAATGAACTCACTGATGTT 1447  
Db 309 TGTCTCAAGTAACTTACCAATTTACCCAAATGAAATGAACTCACTGATGTT 340  
Qy 1448 CAGTCTTCGAAATCAGAGCTGAAAGTGAATCGAATGTGCTGAAACACCGTGTAGGAA 1507  
Db 341 -----AGGAA 346  
Qy 1508 TCGTGTACAGATCTTACATGAGACGTGCTCTCGTGAAGGATCTTCAACGCGGAGGCG 1567  
Db 347 TCGTGTACAGATCTTACATGAGACGTGCTCTCGTGAAGGATCTTCAACGCGGAGGCG 406  
Qy 1568 AGCTGTGAGGAGACAGAGGAGAGCGGAGTTGAGTTCGCTCCCAAGACCTGAGGG 1627  
Db 407 AGCTGTGAGGAGAGAGAGGAGAGCGGAGTTGAGTTCGCTCCCAAGACCTGAGGG 466  
Qy 1628 ATTTGAGCGCCATTTGTTTCAGAGAGTACTCCCTGAAGCTGCGGTATATCTGAGCCAG 1687  
Db 467 ATTTGAGCGCCATTTGTTTCAGAGAGTACTCCCTGAAGCTGCGGTATATCTGAGCCAG 526  
Qy 1688 CATCAAGCAGGAAAGTTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1747  
Db 527 CATCAAGCAGGAAAGTTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 555  
Qy 1748 AACATGAGCATTTCAACCTGAGACACAGAGCTTACCTTCCGATTCAGGAACCTTTACATG 1807  
Db 556 -----CAGGAACCTTTACATG 570  
Qy 1808 AGGATGAGCTGGACTTCCATCTGCAAGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTT 1867  
Db 571 AGGATGAGCTGGACTTCCATCTGCAAGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTT 630  
Qy 1868 CCGGATCTCCCGAGAAAGCTTCCGCGAGGCTTCGATGCGGCCCAACATCATCTGTCAG 1927  
Db 631 CCAGATCTCCCGAGAAAGCTTCCGCGAGGCTTCGATGCGGCCCAACATCATCTGTCAG 690  
Qy 1928 CTGCGGTTTATCGACCGGCTTGGCGGCATCAGAGGTTCTTCCACGTCGGTCCAGAGGCC 1987  
Db 691 CTGCGGTTTATCGACCGGCTTGGCGGCATCAGAGGTTCTTCCACGTCGGTCCAGAGGCC 750  
Qy 1988 CTCTAGCGCAGAGCATCAAGCTCGTGGACGAGTTTACCTTACAGCGTGTATCCCGGAGG 2047

751	Db	CTCTAGCGACAGCATCAAGCTCGTGGACGAGTTTCACTTACAGCGTGAATCCGCGCGAGG	810
2048	Qy	AAGGCCGAGATCGTGGAGCCCGCGGCCAGCGCAACAGGAGAGGTACGTGCACATGAC	2107
811	Db	AAGGCCGAGATCGTGGAGTCCGGCGACGCGCAACAGGAGA-----	853
2108	Qy	TGTTTCGATTCTTCACTTTCATCGTCTTTGGCCGGGATGGACCTGTATCTCTGATTGATTATAT	2167
854	Db	-----	853
2168	Qy	ATCCGTGTGACTTGTGAGGACAAATTTAAATGGGCGAGATGAAGACGACATCCTGTCAAG	2227
854	Db	-----AGATGAAGCACGACATCCTGTCAAG	878
2228	Qy	GTTTCATCGACTAGGCGAGCGCGCGACGACGCGCGCTTTCGGGAGCGACAAGAGCT	2287
879	Db	GTTTCATCGAGCTGGGCGAGCGCGGACGACGCGCGCTTTCGGGAGCGATAGAGCCT	938
2288	Qy	CCGGGACGTGGTGTCTCAACTTTCGTGATCGCCGGCGGGACACGACGGCGACGCTGTC	2347
939	Db	CCGGGACGTGGTGTCTCAACTTTCGTGATCGCCGGCGGGACACGACGGCGACGCTGTC	998
2348	Qy	GTGGTTACGCACATGGCCATGTCCACCCGAGCTGGCCGCGAGAGCTCGCCGCGAGCT	2407
999	Db	GTGGTTACGCACATGGCCATGTCCACCCGAGCTGGCCGCGAGAGCTCGCCGCGAGCT	1058
2408	Qy	GTGGCGTTTCGAGCGGAGCGCGCGGAGGAGGCGCTCGCGCTGCTGCCTCGCGCGG	2467
1059	Db	GTGGCGGTTTCGAGCGGAGCGCGCGGAGGAGGCGCTCACCTCGTGTCTGCGCGG	1118
2468	Qy	CGTTGACGCGCGACGACAAGCGTTTCGCGCGCCCGCTGGCGCAGTTTCGCGGGCTCTCAC	2527
1119	Db	CGTTGACGCGCGACGACAAGCGTTTCGCGCGCCCGCTGGCGCAGTTTCGCGGGCTCTCAC	1178
2528	Qy	CTACGACAGCCTCGGCAAGCTGGTCTTACTCTCAAGCTTGGTCAACGAGACGCTCGCCT	2587
1179	Db	CTACGACAGCCTCGGCAAGCTGGTCTTACTCTCAAGCTTGGTCAACGAGACGCTCGCCT	1238
2588	Qy	GTACCCCGCGTCCCTCAGGTGAGCGCGCCGACACGCGACCTCGGTCCAGAGCACAGC	2647
1239	Db	GTACCCCGCGTCCCTCAGGTGAGCGCGCCGACACGCGACCTCGGTCCAGAGCACAGC	1254
2648	Qy	ATGCAGTGAAGTGACCTGAATGCATATGCACATGCACCTTGC CGCGCGAGACCCCAAGG	2707
1255	Db	-----CAGGACCCCAAGG	1267
2708	Qy	GGATCTCGAGGACGACGTGTCGCGACGCGGACGAGGTGAGGCGCGCGGATGGTGA	2767
1268	Db	GGATCTCGAGGACGACGTGTCGCGACGCGGACGAGGTGAGGCGCGCGGATGGTGA	1327
2768	Qy	CGTACGTGCCCTACTCCGATGGGCGGATGGAGTACAACTGGGCGCCGACGCGCGAGCT	2827
1328	Db	CGTACGTGCCCTACTCCGATGGGCGGATGGAGTACAACTGGGCGCCGACGCGCGAGCT	1387
2828	Qy	TCCGCGCGAGCGGTGGATCAACAGGATGGCGGTTCCGCAACGCGTCCCGGTTCAAAGT	2887
1388	Db	TCCGCGCGAGCGGTGGATCAACAGGATGGCGGTTCCGCAACGCGTCCCGGTTCAAAGT	1447
2888	Qy	TCAAGCGGTTCCAGCGGGGCGGAGGATCTGCTTGGGCAAGGACTCGGCGTACTTGAGA	2947
1448	Db	TCAAGCGGTTCCAGCGGGGCGGAGGATCTGCTTGGGCAAGGACTCGGCGTACTTGAGA	1507
2948	Qy	TGAAGATCGCGTGGCCATCCTCTTGCGCTTCTACGCTTCGCGTCTGAGGGGGCACC	3007
1508	Db	TGAAGATCGCGTGGCCATCCTCTTGCGCTTCTACGCTTCGCGTCTGAGGGGGCACC	1567
3008	Qy	CGGTGACGTACCGCATGATGACCATCTCTCCATGGCGCAGCGCTCAAGGTCCGCGTCT	3067
1568	Db	CGGTGACGTACCGCATGATGACCATCTCTCCATGGCGCAGCGCTCAAGGTCCGCGTCT	1627
3068	Qy	CTTAGGCCGCTGTGATGTTCATGGCGATTG-----GGATATCATCCCGTTAATCC-----	3117

Db	1628	CTAGGCGGCTCATGTCTATGGCCATTTTGGATATGATATCGTCCCGCTTAATCCACGAC	168
Qy	3118	-----TTAAAAATTGTCATGCATGTAAGGAAAGCGATGGGTTTCAT	3163
Db	1688	AAATAACGCTCGTGTGTACAAATTTGTCATGCATGTAAGGAAAGCGATGGGTTTCAT	1747
Qy	3164	TGCTGGCTTGGCTTAAGCCCTTAAAACTCGTGGGTCCTTGGCAACACCATCATCTAG	3223
Db	1748	TGCTGGCTTGGCTTAAGCCCTTAAAACTCGTGGGTCCTTGGCAACACCATCATCTAG	1807
Qy	3224	TGTTTTGTACTTACTTACTCTCAGTGGAGTGTAGTGACGAGCATACAAGTTTCATCATATATA	3283
Db	1808	TGTTTTGTACTTACTTACTCTCAGTGGAGTGTAGTGACGAGCATACAAGTTTCATCATATATA	1867
Qy	3284	TTATCCTCTTTCTT 3297	
Db	1868	TTATCCTCTTTCTT 1881	
RESULT 4			
LOCUS	AC073556	160738 bp	DNA linear
DEFINITION	Oryza sativa chromosome 3 BAC OSUNBa0091P11 genomic sequence, complete sequence.		
ACCESSION	AC073556		
VERSION	AC073556.11	GI:18997259	
KEYWORDS	HTG.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE	Gansberger, K., Brenner, M., Burgess, S., Hance, M., Shvartsbeyn, M., Tsitrin, T., Riggs, F., Hsiao, J., Zismann, V., Blunt, S., Pai, G., VanAken, S.E., Uterback, T.R., Feldblyum, T.V., Kalb, E., Quackenbush, J., Salzberg, S.L., White, O., and Fraser, C.M.		
AUTHORS	Oryza sativa chromosome 3 BAC OSUNBa0091P11 genomic sequence		
TITLE	Unpublished		
JOURNAL	2 (bases 1 to 160738)		
REFERENCE	Buell, R.		
AUTHORS	Buell, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
REFERENCE	3 (bases 1 to 160738)		
AUTHORS	Buell, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
REFERENCE	4 (bases 1 to 160738)		
AUTHORS	Buell, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
REFERENCE	5 (bases 1 to 160738)		
AUTHORS	Buell, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-MAR-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org		
REFERENCE	6 (bases 1 to 160738)		
AUTHORS	Buell, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-APR-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org		
COMMENT	On Feb 28, 2002 this sequence version replaced gi:18921330. Address all correspondence to:rice@tigr.org		
	BAC clone OSUNBa0091P11 is from Oryza sativa chromosome 3		
	The orientation of the sequence is from SP6 to T7 end of the BAC clone.		
	Genes were identified by a combination of several methods: Gene prediction programs including Fgenesh ( <a href="http://www.softberry.com/">http://www.softberry.com/</a> ),		

genscan and Genscan+ (Chris Burge, <http://CCR-081.mt.edu/GENSCAN.html>), GeneMarkEMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer (Mihaela Petera and Steven Salzberg, [contact mpertea@igrt.org](http://contact.mpertea@igrt.org)), searches of the complete sequence against a peptide database and the plant EST database at tigr (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are annotated by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

## FEATURES

source

1. .160738

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosomes="3"

/map="SL13095"

/clone="OSJNBa0091P11"

/notes="japonica cultivar-group"

1. .2006

/gene="OSJNBa0091P11.1"

/notes="near identical to sucrose transporter (C terminus)"

GB:BA24071 GI:2723471 (Oryza sativa) (Plant Cell Physiol

38 (12), 1389-1396 (1997)); EST C27284, D40515, A0108830

from this gene"

join(&lt;1. .41,267. .461,580. .723,833. .1066,1169. .1232,

1316. .1359,1476. .2006)

/gene="OSJNBa0091P11.1"

join(&lt;3. .41,267. .461,580. .723,833. .1066,1169. .1232,

1316. .1359,1476. .1649)

/gene="OSJNBa0091P11.1"

/codon\_start=1

/product="sucrose transporter, 5'-partial"

/protein\_id="AAL84308.1"

/db\_xref="GI:19071641"

/translation="ACANLKAGFLVAVIFLSCLVITLIFAKEVPPKGNALPTKNE

PAEPTGTLAVLKGFRNPTGMPSVLIVTGLTSLWFFPFLYDMDGRIYHGDPK

GDPOQTEAFNOGVKAGAFGLLINSIVLGSFLIEPMCKVKGVVWVTSNFIACI

AATPLSFNSLKDFTVOKAITADKSIKAVCLVLEFLGVPLAVLSVPPFAVTAOLA

ATFGGQGLCTGVNLISIVTPOVIALGAGPWFELFGKNIPAFGLASGFALLGGVAG

IFLLPKIKRQFRSVSMGGH"

1662. .1714

/rpt\_family="(CGG)n"

complement(2364. .2386)

/rpt\_family="(GA)n"

complement(3213. .3250)

/rpt\_family="AT-rich"

complement(3415. .3448)

/rpt\_family="AT-rich"

complement(4151. .5119)

/gene="OSJNBa0091P11.1"

/notes="predicted by fgenesh"

complement(&lt;4151. .5119)

/gene="OSJNBa0091P11.1"

complement(4151. .5119)

/gene="OSJNBa0091P11.1"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAL84309.1"

/db\_xref="GI:19071642"

/translation="MGCTTGRQARLDHRCPSLALPRCOSFPARCADGAVHVRLT

STTLGSLVDKGAFAAEAFMRVPTTMTTPNPEPAIDAWALMAGLEHSPLV

PPFAHGSFPITAVPELAASAKVTPPLVPEKKKASPVAPRKAVLYFTSARGVRA

THECCCLARALGGVVRVDYRSMRGRDELHGLGRGAALAKWAPAAPAL

PSLVFDGELVGNDELKRLHLAGELARLAGESAAFGAAGACEACADVRFVLGCAC

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

SGSKYVVDGDDDDNPLDGGGGGPRCTECNENGIVRCPVCC"

complement(4245. .4267)

/rpt\_family="(CGA)n"

complement(4337. .4365)

/rpt\_family="GC-rich"

complement(4468. .4491)

/rpt\_family="GC-rich"

complement(4925. .4955)

/rpt\_family="GC-rich"

complement(7403. .7423)

/rpt\_family="AT-rich"

complement(7998. .8021)

/rpt\_family="AT-rich"

8348. .8398

/rpt\_family="(AGGGGG)n"

complement(8923. .8951)

/rpt\_family="(GA)n"

complement(8983. .9010)

/rpt\_family="AT-rich"

9203. .9268

/rpt\_family="tandem"

11311. .12015

/gene="OSJNBa0091P11.18"

/note="predicted by fgenesh"

join(&lt;11311. .11749,11879. .&gt;12015)

/gene="OSJNBa0091P11.18"

join(11311. .11743,11879. .12015)

/gene="OSJNBa0091P11.18"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAL84310.1"

/db\_xref="GI:13071643"

/translation="MGGLGVLDLDFARALRLWLFWECPDKPWGTTPHVDELDE

QIFVAKVTICNGQAKFWTSNWIHQPLKYLALPNHSGKRLVQDALTNDKMI

EDIRHLSMLPIREFFAVRLWNETNLEEGVEDITITWRWTNTGIGHPTACNSDNGP

ITTFANSTATGWTKNYISRIVHSREKGTG"

complement(12707. .12726)

/rpt\_family="(GA)n"

12726. .12754

/rpt\_family="tandem"

12831. .15245

/gene="OSJNBa0091P11.3"

/note="similar to DNA binding protein GB:BAA05625

GI:1435022 (Daucus carota); EST D48120 from this gene"

join(&lt;12831. .13484,14329. .15245)

/gene="OSJNBa0091P11.3"

join(12831. .13484,14329. .14775)

/gene="OSJNBa0091P11.3"

/codon\_start=1

/product="putative DNA-binding protein"

/protein\_id="AAL84311.1"

/db\_xref="GI:19071644"

/translation="MASNGMASPSFFPFPNELLMAQQAAAPPHDQEHHHHHHGH

HGHHEHQQQQHHHHLGPPPPPPHPHPNPLFPSSAQCSLOEFQFMALWPKRPMY

GDGGGGDEVGGEDELSDGSGAGEKRRINVEQVTLKFNGLKLEPRKQVL

ARALGLQPROVALWFORRARKWTKOLEXDALKQLDVAENDALNNHKKLQAE

IYALKGREASSELINLKETASCSNSSEINIDISRTTPDAAALDTAPTAGHH

HGGGGGGGGGGMIFPYSIARPSAGSGGVDDIQLLHSSGGGAGGPKMEHGGGNNVQ

AASVPTASFGNLLCGVDEPPFPFPWPDHGHF"

complement(12936. .13011)

/rpt\_family="(TGG)n"

13136. .13179

/rpt\_family="(CGG)n"

complement(13641. .13662)

/rpt\_family="(CAG)n"

13848. .13884

/rpt\_family="(CGAT)n"

14509. .14541

/rpt\_family="tandem"

complement(15170. .15194)

/rpt\_family="AT-rich"

complement(16201. .16222)

/rpt\_family="(TA)n"

repeat_region	complement(18417..18442)	/rpt_family="AT_rich"
Query Match	28.6%;	Score 1114.8; DB 8; Length 160738;
Best Local Similarity	73.3%;	Pred. No. 1.3e-158;
Matches 1671; Conservative	0;	Mismatches 432; Indels 177; Gaps 12;
Qy	836	TTTTCTTGTTGTCAGATCACAAAAAGGAACACAAAGGCTAAGCCTCTTACTTGTTCGGG
Db	95440	TTTTCTAGTGGACAGATCACAAAAAGGAACAACTAAAG--CTAAGATCCAACCTCTTAAGGG
Qy	896	AGTTAGTTCAGGACACCATATGATGAAGAATCTTAATTTGGGTCACACCAAGATT
Db	95498	TGTTAGGTTAGGGACACCATATGATGAGA-CAATCTTTAATTTTGGTCAACAAAGATT
Qy	956	GTCTCTCTCGAGTTGGGGGTCCCTAAGGTTGGTAGCAATAACCCAAATATATCACT
Db	95557	GTCTC-----AAGTTGGTAGCATCAGTGGCCAAATATATCACT
Qy	1016	AACAAACCCATCCATGCTACATACATATAGCATCCATCATCTTGTAGACTGGACCTT
Db	95596	AA-----CTATGCCATCCAAATGCTACATAGCATCTCTTGTAGACTGAACCTT
Qy	1076	CATCAAGACCACTGGAGGAAGCTCATATCAGCGCGGGAGCCCATCGCCATCTTCTCC
Db	95646	CATGAAGAGCCCATGGAGGAAGCTCATGCAATGCCAGTGA-----CATCATCTTCTCC
Qy	1136	ACTAGCAGGCGCTCAACAGTATATCGGCTCTCTGTTGTCTCTCATGATCCTGGT
Db	95700	AGTAGCAGGAATCCACAGCTCATAGCTATCTTCTTGTGTCTCTCATGATCTTGGT
Qy	1196	CCAGAGTGGACCTGAGGAAGCAGAAAGCCCGAGATCATGGCCAGTCATCGGTGCAAC
Db	95760	CCACAAGTGGAGCTGAGGAACAGAAAGGGCCAAAGATCATGGCCAATCATCGGCGCAC
Qy	1256	GGTGGAGCAGCTGAGGAATCAACCGGATGCACGACTGGCTTGTTCGGGTACTCTGTCAG
Db	95820	AGTGGAGCAACTGAAGAACTACACAGGATGCTAGTCTGGCTTGTTCGAGTATCTTGTGAA
Qy	1316	GCACAGACAGTGAACGCTTCGACATGCGTTCACTTCTTACACTATACCTGCTGACCCGGT
Db	95880	GGACAGACGGTGAACGCTGACATGCTTTCACCTCTTACACTTACCTTTCGCGACCCGGT
Qy	1376	GAATGTCGAGCATCTCTCAAGACTAACTTCAACCAATTACCCAAAGGTAAATGACCTGAA
Db	95940	GAACTGCGAGCATGTCTGAAGACCAACTTCAACCAATTACCCAAAGGTAAAGAACCATTA
Qy	1436	CTCACTGATGTTCACTCTTCGGAATCAGAGC-----TGAAAGCTGAATCGAATGTGCC
Db	96000	GGATCTTTCAGTGTACTGTAAATGTGCTTGCACAGTACTAAACACTGACACAAAATGT
Qy	1490	TGAACACCGTGTAGGGAATCGTGTACAGATCCTTACATGGACGTGCTCCTCGGTGACGGCA
Db	96060	CTGAAATATGACGGTGAAGTGTACAGGCTTACATGGATGTGCTGCTCGGTGATGGCA
Qy	1550	TCTTCAACCGCAGCGCGAGCTGTGAGGAAGCAGAGGAAGACGGGAGTTTCGAGTTTCG
Db	96120	TATTCATATCCGACGGCGAGATGTGGAGGAAGCAAAAGGAAGACGGCGAGCTTCGAGTTTG
Qy	1610	CCTCCAGAACCTGAGGATTTTCAGCGCCATTTGTTTCAGAGAGTACTCCTCTGAAGCTGT
Db	96180	CCTCCAGAACCTTGAGAGACTTCAGCACTGTGTTTCAGGGAGTACTCCTCTGAAGCTAT
Qy	1670	CGGTTACTAGCCAGGCATCCAAAGCCAGGCAAAAGTTGTGGACATCAGAGGTGAGATCAC
Db	96240	CAAGCAITCTGAGCCAAAGCATGCAAGCCGCGCAGAGTTGTAGACATGCAAGTAAACCACT
Qy	1730	TGCTCCCTTGGCATTGGCCAACTAGAGCAITTCACCTCTGAGACACGAGAGCTACTTGGCG
Db	96300	GAATTCCTTGGCTAATACCTTAACATTTCTTGAAGAAACCAAAATGTTTCAGAAATCTGATG
Qy	1790	-----ATTCAAGAACTTTACATGAGGATGACGCTGGACTCCATCTGCAAG

96360	CAAGAACTAACAAAATTCAGGAATTTGTTTCATGAGAGATGACACTGACTCGATCTGCAAG	96419
1835	GTTGGGTTCCGGGTCGAGATCGGACACGCTGTGCGCGGATCTCCCGAGAACAGCTTCGCG	1894
96420	GTGCGGTTTGGGTTGAGATCGGGAACGCTGTCACTGATCTCCCGGAGAACAGCTTTGGC	96479
1895	CAGCGTTTCGATCGCGCAACATCATCGTCAGCTGCGGTTTCATCGACCCGCTGTGTGCGC	1954
96480	CAGGCAITTCAGCGCTGCCAACATCATCGTCACGCTGGGTTTCATCGATCTCTGTGGCGT	96539
1955	ATCAAGAGGTTCTTCCACGTCGGTCCAGAGGCCCTCTTAGCGCAGAGCATCAAGCTCGTG	2014
96540	CTGAAGAAGTTCTTGGACGTCGGATCAGAGGCTCTCTCTCGAGCAGAGCATGAAGCTGGT	96599
2015	GACGAGTTCACCTACAGCGTGTATCCGCGGAGGAGGCCGAGATCTGCGAGGCCCGGGCC	2074
96600	GATCACTTCACCTACAGCGTGTATCCGCGCGCCGCAAGGCTGAGATCTTTCGAGGCTCAGCC	96659
2075	AGCGGCAAAAGGAGAGAGGTACGTGCACATGACTGTTTTCGATTCTTTCAGTTTCATCGTCT	2134
96660	AGCGGCAAGCAGAGAGAGGTGATCCTTCTCTCTTGTCTCAAGAAATCAGTAGAACTGAAC	96719
2135	GGCGGGATGGAACCTGATCTCTGATTGATTATATATATATATATATATATATATATATAT	2190
96720	TGACATGGTAAATGGTGTGATGATCAGATCGGAAAGGTTTTTGTTCCTGATATATCGTTGAT	96779
2191	ATTAATAATGGCAGATGAAGCAGACATCTCTGTCAACGTTTCATCGAGCTAGCGGAGCGCG	2250
96780	GTAATGGCGAGCAGATCAAGCAGACATATCTGTTCGCGGTTTCATCGAGCTCGGGAGCGCG	96839
2251	CGCAGCAC-----GGCGCGGCTTCGGGACAGCAAGAGAGCTTCGGGACGTTGGTGTCTCA	2304
96840	CGCGGACAGGGGGGCGGAGCTTCGGGAGCAGCAAGAGAGCTTCGCGAGCGTGGTGTCTCA	96899
2305	ACTTCGTGATCGCGGCGGGACACGACGCGACGACGCTGCTGCTGGTTTCACGCAATGG	2364
96900	ACTTCGTGATCGCGCGGCTGCACGACGCGGACGACGCTGCTGCTGGTTTCACGTAATGG	96959
2365	CCATGTCCTCCACCCCGGACGTGSCCGAGAAAGCTGCGCCGACGCTGTGCGCTTCAGGCGCG	2424
96960	CGATGACGACCCCGCGCTGCGCGCAAGCTCCGCGCGAGCTGGCGCGCTTCGAGGATG	97019
2425	AGCGCGCGCGGAGAGGGGCTGCGCTGCTGCTGCGCGCGGCTGACCGCGGACGACA	2484
97020	AGCGCGCGCGGAGAGGGGCTGCGCT-----CGCGGACGCGCGCGCGGCGAGG	97067
2485	AGGCGTTTCGCGCGCGCTGCGCAGTTTCGCGGCGCTCCTTCACCTACGACAGAGCTTCGGCA	2544
97068	CGTCTTCGCGCGCGCTGCGCGAGTTTCGCTGCTGCTGAGCTACGACGCGGTGGGA	97127
2545	AGCTGGTCTTACTCCACGCTGCGTCAACGAGAGCGCTTCGGCTGTACCCCGCGTCCCTC	2604
97128	AGCTGGTGTACTGACAGCGTTCGCTGACGGAGAGCGCTTCGCGCTCTACCCGCGGTGC---	97184
2605	AGGTGAGCGCGCGCACACGCGAGCTTCGCTCCAGAGCAGACATGCAAGTGGAGCT	2664
97185	-----	97184
2665	GAATGCAATGCAATGCACTTTCGCGCGCGCAGAGACCCCAAGGGGATCTCTGGAGGACGAC	2724
97185	-----CGCAGGACCCCAAGGGGATCGTGGAGGACGAC	97216
2725	GTGCTCCGCGGAGGACGAAGGTGAGGCGCGCGGAGTGGTGAAGTACGTACGTGCCCTACTCG	2784
97217	GTGCTCCCGCGAGGACCAAGGTGCGCGCGCGCGGAGTGGTGAAGTACGTACGTGCCCTACTCC	97276
2785	ATGGGCGGATGGAAGTCAAACTGGGGGCCCCGACGCGCGGAGCTTCGCGCGGAGCGGTGG	2844
97277	ATGGGAGGATGGAAGTCAAACTGGGGCCCCGACGCGGCGAGCTTCGCGCGGAGCGGTGG	97336
2845	ATCA---ACGAGGATGGCGGCTTCGCAACGCTGCGCGGTTCAAGTTTCAAGGCTTCAG	2901
97337	CTCAGCGGCGAGCGCGCGCTTCGGAACGCGTTCGCGGTTCAAGTTTCAAGGCTTCAG	97396



QY 2902 GCGGGCCGAGGATCTGCTGGGCAAGGACTCGGCTACTCGAGATGAAGATGCGCTG 2961  
DB 97397 GCGGGCCGCGGATCTGCTGGGCAAGGACTCGGCTACTCGAGATGAAGATGCGCTG 97456  
QY 2962 GCGATCTCTTGGCTCTACAGCTTCGGCTGCTGGGCGGACCGGTCGAGTACCGC 3021  
DB 97457 GCGATCTCTTGGCTCTACAGCTTCGGCTGCTGGGCGGACCGGTCGAGTACCGC 97516  
QY 3022 ATGATGACCATCTCTCCATGGGCGGCTCAAGGTCGGCTCTCTAGGGCGGCTCTGA 3081  
DB 97517 ATGATGACCATCTCTCCATGGGCTCAAGGTCGGCTCTCTAGGGCGGCTCTGA 97576

RESULT 5  
AX453576  
LOCUS AX453576 1092 bp DNA linear PAT 02-SEP-2002  
DEFINITION Sequence 5 from Patent WO0226789.  
ACCESSION AX453576  
VERSION AX453576.1 GI:21712818  
KEYWORDS  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1  
AUTHORS Albertsen, M.C., Fox, T., Huffman, G. and Trimmell, M.  
TITLE Nucleotide sequences mediating male fertility and method of using  
same  
JOURNAL Patent: WO 0226789-A 5 04-APR-2002;  
PIONEER HI-BRED INTERNATIONAL, INC. (US)  
FEATURES  
source  
1. .1092  
/organism="Zea mays"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4577"

ORIGIN  
Query Match 28.0%; Score 1092; DB 6; Length 1092;  
Best Local Similarity 100.0%; Pred. No. 5.1e-155;  
Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAATTCACAGCGAGCCCTTGTAGCAGAGATGTTGTGATGCAAGTCGGCGGAAATGAGT 60  
DB 1 GAATTCACAGCGAGCCCTTGTAGCAGAGATGTTGTGATGCAAGTCGGCGGAAATGAGT 60  
QY 61 GCGTGCTGAGAGCAACGCTGAGGGGTTCCAGGGATGCAATGGCTATGGCAATCGGCTAG 120  
DB 61 GCGTGCTGAGAGCAACGCTGAGGGGTTCCAGGGATGCAATGGCTATGGCAATCGGCTAG 120  
QY 121 AGGTGGAGGACAAGGTGCTGAGGATGGGAGGCAACCTATGGCAAGTTCGTGAAGAGGC 180  
DB 121 AGGTGGAGGACAAGGTGCTGAGGATGGGAGGCAACCTATGGCAAGTTCGTGAAGAGGC 180  
QY 181 ACGCAATGAGAGATCTATTGAGATTACAGTGGATGCGGCAACAAATTCACCTTTAGA 240  
DB 181 ACGCAATGAGAGATCTATTGAGATTACAGTGGATGCGGCAACAAATTCACCTTTAGA 240  
QY 241 TTTTGTACTGTCACTCTACTTTATTCCTTGGTGGCACTTCCCAATAGGCTCATGTT 300  
DB 241 TTTTGTACTGTCACTCTACTTTATTCCTTGGTGGCACTTCCCAATAGGCTCATGTT 300  
QY 301 AATCAATGATTAGTATTATTCAGCAATATCTTGTGTTGTTGACATTTATATATATGTG 360  
DB 301 AATCAATGATTAGTATTATTCAGCAATATCTTGTGTTGTTGACATTTATATATATGTG 360  
QY 361 GGGTGAGCGGATTAAATATCATTCATGAGAGCTTTATCTTCATGCTCTCTGTTTGG 420  
DB 361 GGGTGAGCGGATTAAATATCATTCATGAGAGCTTTATCTTCATGCTCTCTGTTTGG 420  
QY 421 TTTTCAGATCATCTTTCAGTGTTCACAGAAATTTCTCAGTTTCGTCATGTAATTTTG 480

DB 421 TTTTCAGATCATCTTTCAGTGTTCACAGAAATTTTCTCAGTTTGGTCCATGTAATTTTG 480  
QY 481 AAGTCAGGTTTCCTTAAATTCATTAATGCTTCTCTTCTTTCTAGACTAGCAACTGATGA 540  
DB 481 AAGTCAGGTTTCCTTAAATTCATTAATGCTTCTCTTCTTTCTAGACTAGCAACTGATGA 540  
QY 541 CTTTTCACCTTTGGGTTCCAAATTCAGTTCACACAGAAACAAATTCACITTTGGGTTTCA 600  
DB 541 CTTTTCACCTTTGGGTTCCAAATTCAGTTCACACAGAAACAAATTCACITTTGGGTTTCA 600  
QY 601 AATTCCTCTTCAGGATGTACTTTTTCACITTTGAATCTGTCTATGAGAAACAAGAAATGGT 660  
DB 601 AATTCCTCTTCAGGATGTACTTTTTCACITTTGAATCTGTCTATGAGAAACAAGAAATGGT 660  
QY 661 CAGTTTTTAAAGGAAACAATGTACAGATTTTCATTTCAAGACTCTTTCTGGTTGGTGGTTT 720  
DB 661 CAGTTTTTAAAGGAAACAATGTACAGATTTTCATTTCAAGACTCTTTCTGGTTGGTGGTTT 720  
QY 721 CAGACTTTTTTGTACCAAGCTGATGGATCACAATCTTGTTCCTCAAGTCTGTATAACAGAA 780  
DB 721 CAGACTTTTTTGTACCAAGCTGATGGATCACAATCTTGTTCCTCAAGTCTGTATAACAGAA 780  
QY 841 TTGGTTGGGAGATCACAATAAGGAAACAAAGGCTTAAGGCTTCTTCTGGGAGTTA 900  
DB 841 TTGGTTGGGAGATCACAATAAGGAAACAAAGGCTTAAGGCTTCTTCTGGGAGTTA 900  
QY 901 GGTTCAGGACACCATATGATGAAGAAATCTTAAATTTGGGGTCCACCAAGATTTGCTCTC 960  
DB 901 GGTTCAGGACACCATATGATGAAGAAATCTTAAATTTGGGGTCCACCAAGATTTGCTCTC 960  
QY 961 TCTCGAGGTTGGGGGTCCTTAAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1020  
DB 961 TCTCGAGGTTGGGGGTCCTTAAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1020  
QY 1021 ACCCAATCCATGCTATACATACATAGCATCATGATCCATCCTTTGAGTTCAGTTCAGTTCATCA 1080  
DB 1021 ACCCAATCCATGCTATACATACATAGCATCATGATCCATCCTTTGAGTTCAGTTCAGTTCATCA 1080  
QY 1081 AGAGCACCATGG 1092  
DB 1081 AGAGCACCATGG 1092

RESULT 6  
AF366296  
LOCUS AF366296 1091 bp DNA linear PLN 14-MAY-2001  
DEFINITION Zea mays cytochrome P450-like protein (ms\*-sb200) gene, promoter  
sequence.  
ACCESSION AF366296  
VERSION AF366296.1 GI:14030555  
KEYWORDS  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 1091)  
AUTHORS Fox, T.W., Trimmell, M.R. and Albertsen, M.C.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2001) Trait and Technology Development, Pioneer  
Hi-Bred Intl. Inc., 7300 N.W. 62nd Ave. P.O. Box 1004, Johnston, IA  
50131-1004, USA  
FEATURES  
source  
1. .1091  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4577"  
1. -.1091  
/gene="ms\*-sb200"



```
/note="cytochrome p450-like protein; anther specific"
1. .1088
/promoter
TATA_signal
1006 .1009
/gene="ms*-sb200"
/gene="ms*-sb200"

ORIGIN
Query Match      28.0%; Score 1091; DB 8; Length 1091;
Best Local Similarity 100.0%; Pred. No. 7.2e-155;
Matches 1091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGAAGCAGAGCCCTGTAGCAGAGAGTGTGCTGATGTCAGTTCGGCGGAATGAGT 60
DB 1 GAATTCGAAGCAGAGCCCTGTAGCAGAGAGTGTGCTGATGTCAGTTCGGCGGAATGAGT 60

QY 61 GGTGCTGAGAGCAACGCTGAGGGGTTCAGGGATGGCAATGGCTATGGCAATCGGCTAG 120
DB 61 GGTGCTGAGAGCAACGCTGAGGGGTTCAGGGATGGCAATGGCTATGGCAATCGGCTAG 120

QY 121 AGGTGGAGCACAAGGTGGTGGAGATTGGGAGGCAACCTATGGCAAGTTGGTGAAGAGGC 180
DB 121 AGGTGGAGCACAAGGTGGTGGAGATTGGGAGGCAACCTATGGCAAGTTGGTGAAGAGGC 180

QY 181 AGCAATGAGAGATCTATTCAGACTTACACTGGATGGCGCAACAATTCACCTTTAGA 240
DB 181 AGCAATGAGAGATCTATTCAGACTTACACTGGATGGCGCAACAATTCACCTTTAGA 240

QY 241 TTTTGATCTGTCACCTCTCTTATTCCTGGTGGGCAACTTCCATAGAGCTCATGTT 300
DB 241 TTTTGATCTGTCACCTCTCTTATTCCTGGTGGGCAACTTCCATAGAGCTCATGTT 300

QY 301 AATCAATGATGATGATTAATCAGCAAAATATCTTGTGTTGTCATTAATATGTTG 360
DB 301 AATCAATGATGATGATTAATCAGCAAAATATCTTGTGTTGTCATTAATATGTTG 360

QY 361 GGTGAGAGCGGATTAATATCATCATGAGAGCTTATCTTCATGCTCTCTGATTTGG 420
DB 361 GGTGAGAGCGGATTAATATCATCATGAGAGCTTATCTTCATGCTCTCTGATTTGG 420

QY 421 TTTTCAGATCATTTCTTTCACTGTTTCAAGAAATTTCTCAGTTTGGTCCATGTAATTTTG 480
DB 421 TTTTCAGATCATTTCTTTCACTGTTTCAAGAAATTTCTCAGTTTGGTCCATGTAATTTTG 480

QY 481 AAGTGAAGTTCCTTAATTTCAATATATGCTTCCTTTCTTTCTAGACTAGCACTGATGA 540
DB 481 AAGTGAAGTTCCTTAATTTCAATATATGCTTCCTTTCTTTCTAGACTAGCACTGATGA 540

QY 541 CTTTTCACATTTGGGTTCAAAATTTGACTCACAAGAAACAATTCATCTTTGGGTTTACA 600
DB 541 CTTTTCACATTTGGGTTCAAAATTTGACTCACAAGAAACAATTCATCTTTGGGTTTACA 600

QY 601 AATTCCTCTTCAGGATGTACTTTCACTTGAATGCTCATGTATGAGAAACAAGAAATGGCT 660
DB 601 AATTCCTCTTCAGGATGTACTTTCACTTGAATGCTCATGTATGAGAAACAAGAAATGGCT 660

QY 661 CAGTTTTTAAAGAACATGATACAGATTTCAATTCAGAACTCTTTCTGGTGGTTGATTT 720
DB 661 CAGTTTTTAAAGAACATGATACAGATTTCAATTCAGAACTCTTTCTGGTGGTTGATTT 720

QY 721 CAGACTTTTGTACCAAGCTGATGATCAAAATCTTGTTCCTTCCAAAGTCTGATAACAGAA 780
DB 721 CAGACTTTTGTACCAAGCTGATGATCAAAATCTTGTTCCTTCCAAAGTCTGATAACAGAA 780

QY 781 ACTGGCAACTCTTAATTTGATAAATAAAGAAATAAATAACAGTATCAGATATCTCTTTTC 840
DB 781 ACTGGCAACTCTTAATTTGATAAATAAAGAAATAAATAACAGTATCAGATATCTCTTTTC 840

QY 841 TTGGTGGGAGATCACAAGAAAGAACACAAGAGCTTAAGCTCTTACTTGTTCGGGAGTTA 900
DB 841 TTGGTGGGAGATCACAAGAAAGAACACAAGAGCTTAAGCTCTTACTTGTTCGGGAGTTA 900

QY 901 GGTCAAGGACACCATATGATGAAGAAATCTTAATTTGGGTTCCACCAAGATGTCCTC 960
DB 901 GGTCAAGGACACCATATGATGAAGAAATCTTAATTTGGGTTCCACCAAGATGTCCTC 960

Db 901 GGTCAAGGACACCATATGATGAAGAAATCTTAATTTGGGTTCCACCAAGATGTCCTC 960
Qy 961 TCTCGAGGTTGGGGGGTCCCTTAAGGTTGGTAGTAGCAATACCCCAATATATCCTACCAAA 1020
Db 961 TCTCGAGGTTGGGGGGTCCCTTAAGGTTGGTAGTAGCAATACCCCAATATATCCTACCAAA 1020
Qy 1021 ACCCAATCCATGCTACATACATACATACATACATACATACATACATACATACATACATCA 1080
Db 1021 ACCCAATCCATGCTACATACATACATACATACATACATACATACATACATACATACATCA 1080
Qy 1081 AGAGCACCACATG 1091
Db 1081 AGAGCACCACATG 1091

RESULT 7
AC073178 107214 bp DNA linear PLN 30-JAN-2001
LOCUS Arabidopsis thaliana chromosome 1 BAC F10D13 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC073178
VERSION AC073178.9 GI:12597777
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta;
Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 107214)
AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Maiti, R., Renning, C.M., Koo, H., Fujii, C.Y., Uterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 1 BAC F10D13 genomic sequence
Unpublished
REFERENCE 2 (bases 1 to 107214)
AUTHORS Town, C.D. and Kaul, S.
DIRECT SUBMISSION
JOURNAL Submitted (10-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE 3 (bases 1 to 107214)
AUTHORS Town, C.D. and Kaul, S.
DIRECT SUBMISSION
JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE 4 (bases 1 to 107214)
AUTHORS Town, C.D. and Kaul, S.
DIRECT SUBMISSION
JOURNAL Submitted (30-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 30, 2001 this sequence version replaced gi:12320700.
Address all correspondence to: atc@tigr.org

BAC clone F10D13 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GenemarkRMM (Mark Borodovsky,
http://genemark.biology.gatech.edu/Genemark/), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
Genesplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
```

Simple repeats are identified by repeatmasker (Arian Smit,  
http://ftp.genome.washington.edu/RN/RepeatMasker.html).

## FEATURES

## source

1. .107214

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/chromosome="I"

/map="mi353"

/clone="F10D13"

/complement(877..1232)

/gene="F10D13.1"

/complement(<877..1232)

/gene="F10D13.1"

/complement(<877..1232)

/gene="F10D13.1"

/codon\_start=1

/product="hypothetical protein, 3' partial"

/protein\_id="AAG60090.1"

/db\_xref="GI:1259778"

/translation="MNPDNPDLSDNDSSAWRELTTLTAQSDPFDRTSNILSDFGWNL

HSSDHPGLRFDSDLTQTGVKPTVTWSCSSAAVSVATSTNNPSATSSSSBDP

AENSTASAEKTPPEP"

1247..1303

/rpt\_family="(GA)n"

1311..1355

/rpt\_family="(GAAA)n"

/complement(2071..2094)

/rpt\_family="AT\_rich"

5248..7059

/gene="F10D13.2"

/notes="predicted by genemark hm"

join(<5248..5516,6524..6622,6699..6780,6874..>7059)

/gene="F10D13.2"

join(5248..5516,6524..6622,6699..6780,6874..7059)

/gene="F10D13.2"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAG60092.1"

/db\_xref="GI:12597780"

/translation="MKTNRNPINILIVFLLTARAATNWNTRTHRTVPKVOHAYV

AYHRCSERSRYASMCLEIRHSRQPLFSPPPPEIDQSMMESTVORAKP

EHAAVVDQWETEIEIKRKKYKLEKIVSWEDKKRKAARLKHRTSREVEKTLKA

TQRFDERIEIIVASARAHAYESRIKEELKYKAKANLMTTGRKPESTCL"

complement(7461..8273)

/gene="F10D13.3"

complement(<7461..8273)

/gene="F10D13.3"

complement(7461..8273)

/gene="F10D13.3"

/codon\_start=1

/product="unknown protein"

/protein\_id="AAG60094.1"

/db\_xref="GI:12597782"

/translation="MWGFASNVIGSMGLKSPKDSQAQSCQSDDEVSNISRDEGLE

CPCWESFNIVENVYVLCGHTLCQNCVGLQSAVLRLSSQDRIPIFFVSCPCWQLL

SFIVYKGNLKFPRKFLFWVSLNGDRTSHGSLVDNQQSAPTPRCMS

LGNHSSNNLVARPLLRNQSTDLLPHDHGNSQPSQLFSFKSLIDFISFTSKPFVI

IFLLVIFATPGSIIILALYFLLTILFAPVAGLVLYFAYFILRLVNEITSS"

9624..12967

/gene="F10D13.28"

join(<9624..9944,10045..10122,10256..10311,10628..10747,

10873..10943,11031..11124,11351..11554,11637..11888,

12006..12114,12198..12265,12374..12443,12524..12598,

12713..12804,12892..12967)

/gene="F10D13.28"

join(9624..9944,10045..10122,10256..10311,10628..10747,

10873..10943,11031..11124,11351..11554,11637..11888,

12006..12114,12198..12265,12374..12443,12524..12598,

12713..12804,12892..12967)

/gene="F10D13.28"

/codon\_start=1

/product="unknown protein"

/protein\_id="AAG60116.1"

/db\_xref="GI:12597804"

/translation="MYQTIPTAPTTRGGTPTESGYVVTLDQIPRWSDEQSSLEDE

TGDPHNSPRYANPLASSSEAGSGNMGVSPFVDHEINSRIYLRGEPNLEVDVAV

KSTNENDEAHSSPGLHVAAGPGLAEQCATLGGCTGMAKVTNAYDLPRVHTVGP

KYAVKYHTAAENALSHCVSRCLLELIDGLOSIALGCIYTEAKNYPRPAAHVAITV

RRFLKQDKISAVVCTTSTSDTEIKRLPLYPFRDEHEREVAI SKLPADVGDENG

ETVIDERKIRIQALNKNKPPRSPPTPLRPSDTLLLRNSHLDLYLDPAFMSLLKD

PDERKQEWKTAQAQSGNFVKLIGFDLGGPPLSAAEYLSHRSYLAKANSINLSE

IAEMKIVRGVDTGTEHFVMVVAHFLRLCDLDERFLVYVKEFVPIQKYSIVSE

HSAAIQPOLGMWMLRLEQILGRKHQNLQAIYVHLHTFLKATILTMQFVDNVVW

KKVYADRLLQLFKVVPREQLTIPDFVQHDLEVNGGKGLI VDPRTKVVYQRP"

13205..15568

/gene="F10D13.4"

/note="contains Pfam profile: PF01535: PPR repeat"

<13205..>15568

/gene="F10D13.4"

13205..15568

/gene="F10D13.4"

/codon\_start=1

/product="PPR-repeat protein, putative"

/protein\_id="AAG60097.1"

/db\_xref="GI:12597785"

/translation="MTQYMLFRSCSLKLVSLQHAHLVLTGRLRDLPLVTKLIESY

AFWSPDSRIVPEAFVYPDPFMYGVLIKCNVWCHLLDAAIDLHRLVSETTQISKVF

FFSVLRACAGSREHLSVSGKVHGRILKGVDDDAVLETSLLCMYGTGNLSDAEKVFD

GMFVRDLVAVSTSLCSENGEVVAKRMFKCNVDGVPDVTMLSVVSGCAELGCL

RIARSVHQITRKMFDLDTLCTSLNLTMYSCGDLLSSRIFEKAKNAWTAWIS

SYNRFGESEKALRFSEMIKGIENPLVTLYSVLSCGLIGLIRKSGVHGFAVRRL

DPNYSLSLALVELVAECGLCTVLRVSDRNIVAWNSLISLYAHRGWVIOALGL

FFQWYTORIKPDAAFTLASSISACENAGLVPLGKQIHGVIRTVDSDFVQNSLIDMS

KSGVDSASTVFNQIKHRSVVTWNSMLCFSGQNSVEAISLFDYVHSHYLEMVEYTF

LAVIQACSSIGLEKQWVHKHLIISGLKDLFTDALIDYAKCGDLNAAEVFRFAMS

SRSIVQSSMINAYMGRIGSAITFQNMVBSGTKNVEVFNKCVGLASAGHVSVEEG

KYFNMKISFGVPSNEHFACEIDLLSRDLKEAYRTIKEMPFADASVWGLSVNGC

RTHQMDI IKAJNDLSDIVTDGTGYTLLSNIAEEGEWFERLRISAMKSNLKKV

PGYSATEIDQVFRFGAGENRIQTDIEYRFLGNLQNLNEEHVDS"

complement(15576..18785)

/gene="F10D13.5"

complement(join(<15576..17755,17869..18424,18522..>18785))

/gene="F10D13.5"

complement(join(15843..17755,17869..18424,18522..>18785))

/gene="F10D13.5"

/codon\_start=1

/product="unknown protein"

/protein\_id="AAG60100.1"

/db\_xref="GI:12597788"

/translation="MANRALNCVFGSGMGETVHCGSLSSMRDLSEDISNTCSYMY

CGGDKTLPGYQNGFSARPPTDSYERDFLKTMLHEAVFKQVYELHRLYRTQKSL

MAEVKKNFVDHLNNEPTPGSIGKFLFGNSICGEGSTQDCNCKNKLVEVRPV

KVRTMIDQLPADEYLTHTGNTTCTPPYEQSKVEGENIFPESHRNDSGSSLLMKN

NGFTDLNEPQCQDSVFSSSSRDLYSLYGANISHVQGVWEKNTSQNGWMLAENG

KSTPRDKLCLPSHSQVLSAFQPLGYPTSDHSLSGERASKPCVRQNRNPVSXDS

YVSSVASVNPVLNHCYRPSVPWSHVISSWENRSSSVQKPLQANPFLFTNTQV

RADSSAEMSRSNGNLQGSFSSBSAFNFSVFNHNLNPKGATVNGSLCESVMH

QSLKNQKPKQECSSGLPWIKPKPLKNGKTINGLIDLASANHQPMDERMDGDSNY

VHPQNGRSTVCSNDANLHVEMANSQVRKILGFISQKLSICEHPSLITSSVCIS

FSNPVETKATTLINVEAPLTLESEBGGKPEKDEAGDSVDEILIAEAEEIVTIS

LSYCHRTDEAASSSTDADVKEPSWFTVNTIASCGNDLESKIDACLEARDCECRREC

SGDFEEMATNLNTQTEEDYMPKPLIPEYLKFDGTGMGTISNRPGRQRRRGP

KRDFQDILPLGLASLSRLVEDLQMGFGIMKATGVNWSGMARRSRGRKRLVSN

DRAPVCSIAQPMNNSVQVGLDRLSLTGWGNATRRHRRCFAGTPTPTVLT"

19062..19085

/rpt\_family="(A)n"

19115..19137

/rpt\_family="(GAGAA)n"

complement(19511..19569)

/rpt\_family="(GA)n"

complement(19621..19663)

9.64; Score 375; DB 8; Length 107214;

repeat\_region

repeat\_region

repeat\_region

repeat\_region

Query Match

Best Local Similarity 54.0%; Pred. No. 5.8e-47; Matches 1058; Conservative 0; Mismatches 775; Indels 128; Gaps 9;			
QY	1182	TCATGGATCCCTGGTCCAGAGGTGAGCGCTGAGGAGCAGAGAGGCGCGAGATCATGGCCA	1241
Db	67317	TCATGGATCTCTTACACCGATGGGCAAAAGAACAGAGGTGGTCCAAAGACTTGGCCT	67376
QY	1242	GTCACTCGGTGCAACGGTGGAGCAGCTGAGGAACCTACCAACCGATGCGAGCTGGCTTGT	1301
Db	67377	TTGGTTGGAGCAGCGATCGAACAGTTAACTAATTTGATCGAATGCAATGATGGCTCGT	67436
QY	1302	GGGTACCTGTGACGSCACAGACAGTACCGCTGACATGCGCTTCACTTCTCTACACCTAC	1361
Db	67437	GAGTATCTTTAATCACTCAAGAACAGTAGTGGTCCCAATGCCATTCACACTTATACATAC	67496
QY	1362	ATCGCTGACCGCGTGAATGTGAGCATGTCTCTCAAGACTAACTTCACTTACCCCAAG	1421
Db	67497	ATAGCTGATCCCATCAAGTTGAATATGCTCTGAAACAAACTTCTCCAAATACCCCAAG	67556
QY	1422	GTAAATGACCTGAATCACTCACTGATGTTGAGTCTTCGGAATCAGAGCTGAAGCTGAATCG	1481
Db	67557	GTGAAAAAAGCTCTTTATATCAATATCATGTATATGTTTT-----GTGTGACT	67605
QY	1482	ATGTGCTGACACCGGTGAGGGAATCGGTACAGATCTCTACATGAGCTGCTCTCGG	1541
Db	67606	TTGATTTTTTTGTTGATATATACAGGAGAGACTTATCAITCTTATATGGAAGTATTTGGG	67665
QY	1542	TGACGGCATCTTCAACGGCGAGCTGTGGAGGAAGCAGAGAGAAAGACGGCGAGTTT	1601
Db	67666	AGATGGGATTTTCAATCTGATGAGAGCTTTGGAGGAAACAGAGAAAAACCGCGAGTTT	67725
QY	1602	CGAGTTCGCTCCAGAACTGAGGGAATTCAGCGCCATTTGTTTCAGAGAGTACTCCCT	1661
Db	67726	CGAATTTGCTTCCAGAACTTTAGAGATTTTAGTACTGTAGTGTTTAAAGAGTATAGTCT	67785
QY	1662	GAACTGTGCGGTATACCTGAGCGAGCATCCAAAGCAGCAAGTTGTGGAATCATCCAGT	1721
Db	67786	TAACCTCTTCACTATCTCTTCCCAAGCATCTTTCAAAGCAACAGTAGATATGCAAGT	67845
QY	1722	GAGTCACTGTCTCCCTTGCCATGTCACATGAGCATTTCAACCTTGAGACAGAGCTA	1781
Db	67846	AGA-----ATCTAATTAATCACTAAAGTCTACAACTTAGTTTCTTTACATAACATCT	67900
QY	1782	CCTTCGCGATTCAGGAATTTTACATGAGATGACCGTGGACTCCATCTGCAAGTTGGT	1841
Db	67901	TTTTTTTGGCTCTAGGAATTTGATGAGAACTGACTTTGGATTCAATATATGTAAGTGGAT	67960
QY	1842	TCGGGCTGAGATCGCACGCTGTCGCGGATCTCCCGAGAACAGCTTCGCGCAGGCGT	1901
Db	67961	TTGGTGTGAGATAGGACATTTGGCTCCAGAGTTACAGAGAACTACATTTGTAAGGCTT	68020
QY	1902	TCGATCGCGCAACATCATCTGCTGCGCTTCATCGACCGCTGTGGCGCATCAAGA	1961
Db	68021	TTGATACCGCAATATAATCGTAAACACTTCGTTTCAATAGATCTCTTTGGAAGATGAAA	68080
QY	1962	GGTCTTCACTGCGGTGAGAGCCCTCTAGCGCAGAGATCAAGCTCTGGAGAGT	2021
Db	68081	AGTTCCTTAACATAGGCTCTGAGGCAATGCTTGGTAAAGCATAAAGTAGTCAATGAT	68140
QY	2022	TCACCTTACAGCTGATCCCGGAGGAAGCGCGAGATCTCGAGGCGCGCGGCGAGCGCA	2081
Db	68141	TCACCTTTTCACTATAAGAGAGGAAGAGAGAGTATTAGAGGAC-AAATATCTCT	68199
QY	2082	AACAGAGAGGTPACGTGCACATGACTGTTTCGATCTTTCAGTTCACTCTGTCGCGGG	2141
Db	68200	ACCAACAAACAAACAAACAAACAAACAAAGGTGATAACTTGTGTATACAACTCAAAAATG	68259
QY	2142	ATGGACCTGATCTGTATGATTATATATCTCGGTGTGACTTGTGAGGACAAATTAATGGG	2201
Db	68260	TAAATATATGCACTCTGTCACTTAGTAACATTTTATTTCTCATATGTAATCT-----	68310
QY	2202	CAGATGAAGCAGACATCTCTGTCACCGTTTCATCGAGTAGGCGCGCGAGCAGCGC	2261
Db			
Db	68311	-AGGTGAGCACGATATACTCTCGAGTTTATCGAGAT-----CAGTGACGATCCA	68360
QY	2262	GGCGGCTTCCGGGACGACAGAGCTCCGGGAGCTGGTGTCTCAACTTCGTGATCGCGGG	2321
Db	68361	GATAGCAAGAAACAGAGAAAGCTTAAGAGATATTGCTCTCAACTTTGTAATAGCTGGA	68420
QY	2322	CGGACACAGCGGAGCAGCTGCTGCTGTGTGTACGCACATGGCCATGTCCTCCACCGGAC	2381
Db	68421	AGAGATCAACAGCAACAACTCTCACTTGGCTATATACATGATATATGATGAATGAAAT	68480
QY	2382	GTGCGCAGAGCTGCGCGCGAGCTGTGGCGGTTCGAGCGGAGCGCGCGAGGAG	2441
Db	68481	GTGCTGAGAACTATATCTCAGAGCTACAAGAACTCGAAAAAGAAAGCGGAGCGACT	68540
QY	2442	GGGTGCGGCTGCTGCGCTCGGGCGGCTGACGCCAACACAGCGGTTCGCGCCCGC	2501
Db	68541	AATACATCTGTGATCAAT-----ACGATACAGAGGATTTCAATTCCTTCAACGAGAG	68594
QY	2502	GTGCGCGAGTTTCGCGGCTCTCTCACTACGACAGCTCCGCAAGCTGGTCTACTCTCAC	2561
Db	68595	GTAAAGAAATTCGCGAGACTATTGAATCACTGATTCCTTAGAAACTTCACTACTTACAT	68654
QY	2562	GCCTGCTGTCACCGAGAGCTCCGCTCTACCCCGCTTCCCTCAGGTGAGCGCGCGAC	2621
Db	68655	GCTGTGATCAGAAACACTTCTGCTCTACCTCGCTTCTCTCAGGTAAAGTAAAAATCCA	68714
QY	2622	ACGGAACCTCGGTTCAGAGCAGCATGAGTGAGTGAGTGAGCTGAAATGCAATGCAATGC	2681
Db	68715	TTAAATTTGTGAATTTGAGATATCGAAACGAAAGAAAGAAAAATCGTATATTTCGATAT	68772
QY	2682	ACTTTCGCGCGCGCAGACCCCAAGGGATCTCTGAGAGCAGACCTGCTCCGCGAGCGGAC	2741
Db	68773	-----TCGTGAGGATTCGAAAGAGATATTAGAGATGATATGTTACCAATGGAC	68822
QY	2742	GAAAGTGAAGGCGCGCGGATGTTGAGTACGTACGTCCCTTCTGATGAGGCGCGATGAGTA	2801
Db	68823	AAAAATAAAGCTGAGGAGTGGTCAATATGTTACTTCAATGAGTGGTCTGATGGAATA	68882
QY	2802	CAACTGGGCGCGCAGCGCGAGCTTCCGCGCGGCGGTGGATCAACAGAGGATGGCG	2861
Db	68883	CAATTTGGGATTCAGATGAGCATTTGTTTAAACCCGAAAGATGGCTTA---AAGATGAGT	68939
QY	2862	GTTCGCAACCGCTCGCGCTTCAAGTTTCAAGCTTCCAGCGCTTC-----	2899
Db	68940	CTTTCAAAATGCATCAACCAITCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCA	68999
QY	2900	-----AGCGGCGCGCG 2910	
Db	69000	ACAAGATTTTGTGTTTGAACATGAAGATGATGAGTTTGTGTGAGTGTAGGCTGGACCA	69059
QY	2911	AGGATCTCCCTGGGCAAGACTTCGCGTACCTGAGATGAAGATGGCGCTGCGCATCTC	2970
Db	69060	AGGATATGTTTGGGAAAGGATTCAGCTTATCTGCAAAATGAAGATGGCAATGGCAATCTT	69119
QY	2971	TTGCGCTTCTACAGCTTCCGCTGCTGAGGCGGACCCCGGTGCGATGATGATGATGATGAC	3030
Db	69120	TGCAGATTTCTATAGTTTCAATTTGTTGCAAAATCATCTCTGCAAGTATCGGATGATGAG	69179
QY	3031	ATCTCTCCATGGCGCAGCGCTCAAGTCCGCTCTCTAG 3071	
Db	69180	ATCTTATCCATGGCTCATGGCTTGAAGGTTACTGTATCCAG 69220	
RESULT 8			
AX453574			494 bp DNA linear PAT 12-JUL-2002
LOCUS			Sequence 3 from Patent WO0226789.
DEFINITION			AX453574
ACCESSION			AX453574.1 GI:21712817
VERSION			
KEYWORDS			Sorghum bicolor (sorghum)
SOURCE			Sorghum bicolor
ORGANISM			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;



COMMENT	Medical Center Dr. Rockville, MD 20850, USA, rbuell@tigr.org On Apr 13, 2002 this sequence version replaced gi:1339367. Address all correspondence to:rice@tigr.org	
	BAC clone OSJNBb0005J14 is from Oryza sativa chromosome 10 The orientation of the sequence is from SP6 to T7 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including Fgenesh ( <a href="http://www.softberry.com/">http://www.softberry.com/</a> ), gensecan and Gensecan+ (Chris Burge, <a href="http://CCR-081.mit.edu/GENSECAN.html">http://CCR-081.mit.edu/GENSECAN.html</a> ), GeneMarkHMM (Mark Borodovsky, <a href="http://genemark.biology.gatech.edu/GeneMark/">http://genemark.biology.gatech.edu/GeneMark/</a> ), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR ( <a href="http://www.tigr.org/tdb/cgi.shtml">http://www.tigr.org/tdb/cgi.shtml</a> ). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are annotated as tRNAscan-SE (Sean Eddy, <a href="http://genome.wustl.edu/eddy/tRNAscan-SE/">http://genome.wustl.edu/eddy/tRNAscan-SE/</a> ). Simple repeats are identified by RepeatMasker (Arian Smit, <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> ).	
	This BAC overlaps with rice BAC OSJNBa0096G08 (AC078894) and OSJNBa0034L04 (AC091680).	
FEATURES	This BAC contains 32974 bp chloroplast DNA insertion. Location/Qualifiers 1..116932 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="10" /map="E10477S" /clone="OSJNBb0005J14" /note="japonica cultivar-group" complement(1..2215) /gene="OSJNBb0005J14.5" /note="predicted by fgenesh" complement(join(<1..605,982..1595,1647..1882,2114..>2215)) /gene="OSJNBb0005J14.5" complement(join(<1..605,982..1595,1647..1882,2114..2215)) /gene="OSJNBb0005J14.5" /note="3'-partial" /codon_start=1 /product="hypothetical protein, 3'-partial" /protein_id="AA012486.1" /db_xref="GI:20146750" /translation="MTVLILCNKETSVMIFCLPVPALPGTGIDQAVNASPSSH FGEGHVAQLARGLFQSCPEVTGSDVPSPRYELSVLYPRGGHAGLAPHH FTVAPRTQLIHLRGSSAFROLLRPSGRDTPYPCPYSESQPLTRMVELSEARATQ HDLLEARSLLARYALEARLAEAGVTSVDTASDLGRLCHASHLSHSSRGSRVTP SGKPTSPSPYPSPHPSPYSITPGHTPVASSCGGPRFIRTVRRKVVGSSTVF RDHPARPAPAAAGAQRPVPISSROPIDVSTSEPTTGAFGFEEGRVEPEANHE SANEQSHSNRSHHTASRAVEQPHRSNRNAEVSQSHRSPPAAIHNSREE RIPSPPHVNNHMLMAVOTQILQGLATAIGIQHNAHGHPHTGNRSKLTDFLRS RPEPSQTVPEVPEADWLKDLVERKLNLCVQCTPVEKTLVASHQLRGPAADWENYCNAAH PEPTNIPWDEF" 5249..5434 /rpt_family="(CGG)n" complement(6319..7182) /gene="OSJNBb0005J14.1" /note="similar to glutathione S transferase TSI 1 from this gene"	repeat_region  repeat_region  repeat_region  gene  mRNA  CDS  repeat_region  gene  mRNA  misc_feature  gene  mRNA







mRNA	/locus_tag="OSJNB0005J14.22"	QY	2171	CGTGTGACTTGTGAGGACAAATTTAAATGGCAGATGAAGCAGACGATCCTCTCTCGAGTT	2230
	/note="predicted by fgenesh"		12578	TTGTG-----CAATAGATTTCGAGGAGGATATCTTACAAAGATT	12540
	/complement("<44575..>45057")		2231	CATCGAGCTAGCGCAGGCGCGCGACGACGGCGCGCGCTTCGGGGACGACGAGAGCTCCG	2290
CDS	/locus_tag="OSJNB0005J14.22"	QY	12539	CAT-----CCAGGCAACGACTAGCGATTCTCGGACGGTTGATTACAACTACCTGAG	12489
	/complement("<44575..>45057")		2291	GGAGCTGGTCTCAACTTCGTGATCGCGGGGGGACGACGAGCGGCGACGACGCTGCTG	2350
	/codon_start=1		12488	AGACATCATATTGAACATTGTTCATAGCGGCAAGACCAACAGCGGGTCCGCTTCTG	12429
gene	/translation="WPPLPLPAPLOETPSSSELKPRETIYFSPPTPRSQERRANVKG EDARGDDIVEDIKMATVWRAAKGDPATRTILFIHGLWSIEVDLHLSVFCFGVIF RSHRTFHALYAPHDKKGGERSRSPLPVPRASLLRLPLHPTSRTPSPATHPEK "	QY	2351	GTTACGCGACATGGCCATGTCCTCCACCCCGACGCTGCGCAGAGAGCTGCGCGGAGCTGTG	2410
	/product="hypothetical protein"		12428	GTTCTGTACATGATGTGCAACACCCGGAGTACAGGAGAAGATCTGCCACGAGCCAT	12369
	/protein_id="AAP54715.1"		2411	CGCGTTTCAGGCGGAGCGCGCGCGAGGAGGCGCTGCGCGCTGCTGCGCGGCGCG	2470
Gene	/note="similar to C terminus of glutathione S transferase GB:AAF29773 GI:6856103 (Gossypium hirsutum); EST BE530893 from this gene"	QY	12368	GGAGGCCACCAACGCGCGCGCGCGCTTCCATCGACGAGTTCTCGCAGAGCTGACCGA	12309
	/pseud		2471	TGACGCGGACGACGAGGCGTTGCGCGCGCGCGCTGCGCGAGTTCGCGGGCTCTCACCTA	2530
	/locus_tag="CHR10v09222002.tRNA-Val-3"		12308	CGAGGCACTGAACAGAGATGCATATCTGCACGCTGCACCTGACGAGAGCTCAGGCTATA	12249
tRNA	52388..>52459	QY	2531	CGACAGCTCGGCAAGCTGCTCTACCTCCACGCTCGCTCACCGAGACGCTCCGCGCTGTA	2590
	/locus_tag="CHR10v09222002.tRNA-Val-3"		12248	CCCTGCACTTCCACTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC	12189
			2591	CCCCCGCTCCCTCAGGTGAGCGCGCGCGACGCGCACTCCGCTCCAGAGCAGCAGCATG	2650
Query Match	4.9%; Score 189.2; DB 8; Length 30029;	QY	12188	CTTGCCCAATGCAATGAATCGAAGTATAGTAAACAAAGTAGTGTCA-----	12141
	Best Local Similarity 49.8%; Pred. No. 5.9e-19;		2651	CAGTGAGTGAGACTGMAATGCAATGCATGCATGCTGCGCGCGCGAGAGCCCAAGGGA	2710
	Matches 694; Conservative 0; Mismatches 613; Indels 87; Gaps 5;		12140	-----CTTTGTTTCTGACGATGTTTCAGGATAACAAGCACT	12105
Db		QY	2711	TCCTGGAGACGACGCTGCTCGGAGCGGACGAAAGGTGAGGCGCGCGGATGTTGACCT	2770
			12104	GTTCTCAGACGATGTTATGCCCCACGAGTTCACGTCAGCAAGGGGACATCGTGTCT	12045
			2771	AGTGCCCTTACTGATGGGCGGATGGAGTACAACTGGGGCCCCCGACGCGGAGCTTCC	2830
QY		QY	12044	ACATCCCTACGCGATGGCGGATGGAGAGCTTGTGGGGCAAGACGCTGAACTCTTCC	11985
			2831	GGCGGAGCGGTGATCAACAGAGATGGCGGTTCCGCAACGCGTCCGCTTCAAGTTCA	2890
			11984	GGCCTGAACGTTGGCTCGATGAGACGGCGTCTTTTCAGCAGGAGAGCCGTTCAAAATTTA	11925
Db		QY	2891	CGGCGTTTCCAGGCG 2904	
			11924	CAGCTTTCAGGTG 11911	
RESULT 12		QY			
LOCUS		QY			
DEFINITION		QY			
ACCESSION		QY			
VERSION		QY			
KEYWORDS		QY			
SOURCE		QY			
ORGANISM		QY			
REFERENCE		QY			
AUTHORS		QY			
TITLE		QY			
JOURNAL		QY			
REFERENCE		QY			

AK118803 1695 bp mRNA linear PLN 06-DEC-2002  
Arabidopsis thaliana At1g69500 mRNA for unknown protein, complete  
cds, clone: RAPL21-14-E14.  
AK118803  
AK118803.1 GI:26452619  
VERSION  
FLI CDNA; CAP trapper.  
KEYWORDS  
Arabidopsis thaliana (thale cress)  
SOURCE  
Arabidopsis thaliana  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1  
Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J.,  
Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P.,  
Kawai, J., Hayashizaki, Y. and Shinozaki, K.  
Arabidopsis thaliana full-length cDNA  
Published Only in Database (2002)  
2 (bases 1 to 1695)



**AUTHORS** Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

**TITLE** Direct Submission

**JOURNAL** Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: meki@gsc.riken.go.jp, URL: http://pfgweb.gsc.riken.go.jp, Tel: 81-45-503-9625, Fax: 81-45-503-9586)

**COMMENT** An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified plasmid vector. Please visit our web site (http://pfgweb.gsc.riken.go.jp/) for further details.

**FEATURES** Location/Qualifiers

source 1..1695

organism="Arabidopsis thaliana"

mol\_type="mRNA"

cultivar="Columbia"

db\_xref="taxon:3702"

chromosome="1"

clone="RAF121-14-E14"

note="common name: thale cress"

1..1695

gene /gene="Atlg69500/F10D13\_15"

CDS 110..1546

gene /gene="Atlg69500/F10D13\_15"

codon\_start=1

product="unknown protein"

protein\_id="BAC43393.1"

db\_xref="GI:26452620"

translation="MHDWLVEYLNSRTVVVPMPTTYIADPNINVLKTFNSY PKGVTHSVMEVLGDGIFNSDGLRQKTAPEFASKNLROFSTVVPKYSIKLF TLISASPEQVQDQELLMTLDSICKGVGEIGTLAPLPENHFAKADFNII VTLRFDPMKKFENLGSALLKSIKVNDFYVIRRKAEILLBAQSPPTNNN NNNKVKHDIILSRFEISDDPDSKETERLSRDIIVNFVAGRDITATLWALIMM NENAEKLYSELOLEKESAEATNTLSHQYDTEFNSFKTEFAGLLNYSGLKLH YLHVTITRLYPVPODPKGVLEDDMLPNGTKVAGMVTYVPSYGRMEYNGSD AALFKPERWLKGVFNASPKPFAQAPRICLKDQSAIYLMQMAMAILCRFYKFLH VNPVHKVYRMVILNSAHLKLVTVSRS"

**ORIGIN**

Query Match 4.8%; Score 186; DB 8; Length 1695;

Best Local Similarity 50.7%; Pred. No. 2.6e-18;

Matches 959; Conservative 0; Mismatches 570; Indels 361; Gaps 7;

1182 TCATGGATCCTGGTCCAGAGTGGAGCCCTGAGGAGCAGAAAGCCGAGATCATGGCCA 1241

8 TCATGGATCCTTTTACACCGATGGGACAAAGAAACAGAGTGGTCCAAAGACTTGGCCT 67

1242 GTCATGGTGCACCGTGGAGCAGCTGAGGAATACACCGGATGACGACTGGCTTGTG 1301

68 TTGGTTGAGCAGCGATCGAA CAGTTAACTAATTTGATCGATGATGATGGCTCGTT 127

1302 GGGTACCTGTACGGCAGCAGGACAGTGCACGTCGATCGGTTCACTTCTTACACCTAC 1361

128 GAGTATCTTTTAACTCAAGAACAGTAGTGGTTCCTCAATGCCATTCACACTTATACATAC 187

1362 ATCGTGCACCGGTGAATTCGAGCATGCTTCAAGACTAATTCACCAATTACCCCAAG 1421

188 ATAGCTGATCCCACTCAACGTTGAATATGCTTCGAAACAAACTTCTCCAAATTACCCCAAG 247

1422 GTAAATGACCTGAACCTCACTGATGTTTCAGTCTTCGGAATTCAGAGCTGAAAGCTGAATCG 1481

248 GGAGAGACT----- 256

1482 AATGTCCTGAAACACCGTGTAGGGAATCGGTGACATCTTACATGAGCGTGTCTCTCGG 1541

257 -----TATCATTTCTTATATGGAAGTATTGTTGGG 285

QY 1542 TGACGGCATCTTCAACCCGACGCGAGCTGTGGAGGAGCAGAGGAGGAGCGGAGTTT 1601

DB 286 AGATGGGATTTTCAATTTCTGATGGAGAGCTTTGGAGGAAACAGAGAAAACCGCGAGTTT 345

QY 1602 CGAGTTTCGCTCCAAAGAACCTGAGGGATTTTCAGCGCATTTGTGTTTCAGAGAGTACTCCCT 1661

DB 346 CGAATTTGCTTCCAAAGAACTTTAGAGATTTTAGTACTGTAGTGTGTTTAAAGATAGTACT 405

QY 1662 GAAGCTGTGGGTATATCTGAGCCAGGATCCAGGAGGAGGAGGAGTGTGAGCATGCGAGT 1721

DB 406 TAAACTCTTCACTATCTCTTTCCCAAGCATCTTTCAAAGAGCAACAGTAGATATG----- 460

QY 1722 GAGATCACTGTCTCCCTTGGCCATTCGCCAACATGAGCATTTCAACCTGAGACACGAGAGCTA 1781

DB 461 ----- 460

QY 1782 CTTTCCCGATTCAGAACTTTTACATGAGGATGACGCTGGAGCTCCATCTGCAAGTTGGGT 1841

DB 461 -----CAGGAATTTGTTGATGAGATGACTTTTGGATTCAATATGTAAGTGGAT 509

QY 1842 TCGGGTGCAGATCGGCAACGCTTCGCCGATCTCCCGGAGATCTCCCGGAGAACAGCTTCGCGAGCGT 1901

DB 510 TTGGTGTGAGATAGGAGCATTTGGCTCCAGAGTTTACCAGAGAAATCACTTTGCTTAAGGCTT 569

QY 1902 TCGATCGCGCAACATCATCTGCTACGCTGGGTTTCATCGACCCGCTGTGGCGCATCAAGA 1961

DB 570 TTGATACCGCAATATATCTGTAACATCTGTTTCATAGATCTCTTTTGGAGATGAAA 629

QY 1962 GTTTCCTTCCAGCTCGGTCAGAGCCCTCTTACGCGAGAGCATCAAGCTCTGTGAGCAGT 2021

DB 630 AGTTCTTAAACATAGGCTCTGAGGCATTTGCTTGTAAAGCATAAAGTAGTCAATGATTT 689

QY 2022 TCACCTACAGCGTATCGCGGAGGAGCCGAGATCTGTCGAGGCGCGGCGGCGGCGGCA 2081

DB 690 TCACCTATTACGTCATTAAGAAAGGAGGAGGAGGAGTATTATAGAGCAGC----- 737

QY 2082 AACAGGAGAGTACGTGCATGATGATCTTTTCGATTCTTCAGTTTCATGCTTGTGCGCGG 2141

DB 738 ----- 737

QY 2142 ATGGACCTGATCTCTGATTTATATATATCGGTGATCTGTGAGACAAATTTAAATGGG 2201

DB 738 -----AAATATCTCTTACCAACAAACAAACAAACAAACAAACAAAC 775

QY 2202 CAGATGAAGAGCAGCATCTCTGCTACGGTTTCACTGAGCTAGCGAGGCGCGGAGCAGCGC 2261

DB 776 AAGGTGAAGCAGATATATCTCTGAGGTTTATCGAGAT-----CAGTGACCATCA 826

QY 2262 GCGGCTTTCGGGAGCAGCAAGAGCCTCCGGGACGCTGGTGTCTCACTTCGATGCGCGG 2321

DB 827 GATAGCAAGAAACAGAGAAAGCCTAAGAGATATTTGCTCAACTTTGTTAATAGCTGGA 886

QY 2322 CGGACACGACGGGACGACGCTGTCTGTTGTTTACGACATGCGGATGTCACCCGAGC 2381

DB 887 AGAGATACAACAGCAACAACTCTCACTTTGGGCTATATACATGATTAATGAATGAATAAT 946

QY 2382 GTGCGCGAGAGAGCTGCGCGGAGCTGTGCGGTTTCGAGGCGGAGCGCGCGCGGAGGAG 2441

DB 947 GTTGTGAGAACTATATCTCAGAGCTACAAGAACTCGGAAAGAAAGCGGAGGAGCT 1006

QY 2442 GCGCTGCGGCTGTGTCCTTGGCGGCGCTGACGCGGACGACGAGGCGGCTTCGCGCCCGC 2501

DB 1007 AATACATCTTGCATCAATA-----CGATACAGAGGATTTCAATTCCTTCAACAGAGAAG 1060

QY 2502 GTGGCGAGTGTGCGGGGCTCTCTACCTTACGACAGCTTCGGCAAGCTGCTTACCTCCAC 2561

DB 1061 GTAAACAGAAATTCGAGGAGCTATTGAACCTACGATTCTTTAGGAAACATTCCTACTTACAT 1120

QY 2562 GCTCGGTTCACGAGACGCTCGCTGTACCCGCGCTCCCTCAGTGTAGCGCGCCCGC 2621

DB 1121 GCTGTGATCACAGAAACACTTCTGCTCTACCTCCGCTTCCT----- 1162

QY 2622 ACGGACCTCCGCTCCAGAGCAGCATGTCAGTGTGAGTGCACCTGAATGCAATGCATGC 2681

Db 1163 ----- 1162

QY 2682 ACTTGGCGCGCGAGACCCCAAGGGGATCTGGAGGACGACGTGCTGCGGACGGGAC 2741

Db 1163 -----CAGATCCGAAAGAGTATTAGAGATGATATTTACCAATCGGAC 1209

QY 2742 GAAGGTGAGCGCGCGGAGTGTGACGTACGTGCGCTACTCGATGGCGGAGTGAAGTA 2801

Db 1210 AAAAGTAAAGCTGGAGGATGTCATATGATGACCTTACTCAATGGTCTGATGAATA 1269

QY 2802 CAATCGGGCCCGACGCGCGGAGCTCCGCGCGGAGCGGTGATCAACAGGATCGCGC 2861

Db 1270 CAATCGGGGATCAGATGACGATGTTTAAACCGGAAGATGCGCTTA---AAGATGGAT 1326

QY 2862 GTTCCGCAACCGCTCGCGCTTCAAGTTACGGCGTTCACGGCGGCGGCGGAGGATCTGCGCT 2921

Db 1327 CTTTCAAAATGCAATCACCATTCAAGTTTCAGCTTTCAGCTGGACCAAGGATATGTTT 1386

QY 2922 GGGCAAGGACTCGCGCTACTCGAGATGAAGATGGCGCTGGCGATCTCTTGGCGTTCTA 2981

Db 1387 GCGAAAGGATTCAGCTTATCTGCAAAATGAAGATGGCAATGGCAATCTTTGCGAGATTTCA 1446

QY 2982 CAGCTTCGCGCTCTCGAGGGGCGGCGGACCGGTGACGATGACGATGACGATCTCTCCAT 3041

Db 1447 TAAGTTTCATTGGTGCCAAATCATCTGTCAAAGTATCGGATGATGACGATCTTATCCAT 1506

QY 3042 GCGCGACGGCTCAAGTTCGCGCTCTCTAG 3071

Db 1507 GGTCTGCTGCTTGAAGTTACTGTATCCAG 1536

RESULT 13

AC078894/c 140681 bp DNA linear PLN 20-DEC-2002

LOCUS Oryza sativa chromosome 10 BAC OSJNBa0096G08 genomic sequence,

DEFINITION complete sequence.

ACCESSION AC078894

VERSION AC078894.11 GI:273111268

KEYWORDS HTG.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

REFERENCE 1 (bases 1 to 140681)

AUTHORS Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Gansberger, K., Kim, M.M., Overton II, L.L., Bera, J.J., Tsitrin, T., Krol, M.I., Jarrahi, B.B., Jin, S.S., Koo, H., Zismann, V., Hsiao, J., Blunt, S., Vanaken, S.S., Utterback, T.T., Feldblyum, T.V., Yang, Q.Q., Haas, B.J., Suh, B.B., Peterson, J.J., Quackenbush, J., White, O., Salzberg, S.L. and Fraser, C.M.

TITLE Oryza sativa chromosome 10 BAC OSJNBa0096G08 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (08-AUG-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 3 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (15-NOV-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 4 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (25-APR-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbueller@tigr.org

REFERENCE 5 (bases 1 to 140681)

AUTHORS Buell, R.

TITLE Direct Submission

JOURNAL Submitted (18-DEC-2002) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

6 (bases 1 to 140681)

Buell, R.

Direct Submission

Submitted (20-DEC-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbueller@tigr.org

On Dec 20, 2002 this sequence version replaced gi:27228824.

Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0096G08 is from Oryza sativa chromosome 10

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Fgenesh (<http://www.softberry.com/>), genscan and Genscan+ (Chris Burge, <http://ocr-081.mit.edu/GENSCAN.html>), GeneMarkMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, [contact mpertea@tigr.org](http://contact.mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

Location/Qualifiers

## FEATURES

## source

1. 140681

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="10"

/map="E13638"

/clone="OSJNBa0096G08"

/notes="japonica cultivar-group"

585. 1903

/gene="OSJNBa0096G08.1"

/notes="predicted by fgenesh"

join(<585. 599,1096. 1228,1707. >1903)

/gene="OSJNBa0096G08.1"

join(585. 599,1096. 1228,1707. 1903)

/gene="OSJNBa0096G08.1"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AA00895.1"

/db\_xref="GI:273111269"

/translation="MKYKLCIRGWDPPEDISKYAMIEWFKSQMQEAGIVDREDLY

FTFSDTQLGKLLPQGLSVLAQVLPVLPDPTNSEIKARGFILLASNSVAYSEKRV

WIRTANKFQCT"

5357. 5378

/rpt\_family="(CAG)n"

complement(5695. 5715)

/rpt\_family="AT\_rich"

5882. 5915

/rpt\_family="(CACCT)n"

complement(6798. 6848)

/rpt\_family="AT\_rich"

7098. 9490

/gene="OSJNBa0096G08.2"

/notes="contains Pfam profile PF00514 (Armadillo beta catenin like repeat); EST C71879, AU181935, AU161982, B1806399, AU095513, AU032071 from this gene"

join(7098. 7339,7509. 9490)

/gene="OSJNBa0096G08.2"

join(7330. 7339,7509. 9226)

/gene="OSJNBa0096G08.2"

/codon\_start=1

repeat\_region

repeat\_region

repeat\_region

repeat\_region

gene

mRNA

CDS

```
/product="putative armadillo repeat containing protein"
/protein_id="AA00697.1"
/db_xref="GI:27311271"
/translators="MVSDBVRCSSFDGMSLHCCPAAGPPAPAGTABELLERARSL
VPAALDAARATGGRGKWKVIAARLVRPPCLSDLSHPGCFKNSLCRELLQSVATL
AEALGARGREPRAGKLOMSDLALACKLDLNRDCALLIKTGVLSDATVPPVAP
AAEAADQVRELLARLQIHAENAKHRVNDGLDALREDEKSVLSALGRNVAAVLQV
LTATAPKIREKATVCLLAESGCECLLVSEGLAPPLILRVSGSLVGREKAVITIQ
RLSNPDIARIVGHSVRPLIDICQTDSSISQSAAGALKNSLAVPEVQALABEG
VRVNVLDGCVLGCYAECLQSLTSSNDGLRRVAVSEGLRLSLAYLDGPLPOE
SAVCALNLYSSATSPSLVSLGLVPLRVHVRGSGVGAOAAAAACRVSSSEMKR
LVGSHGCMPLLVLEKASGREVAQAQVAVSLMCLANARDIKKDEKSVNVLQVLE
PSPQNTAKKVAISCLLTSASRCKKLMISHGALGYLKLSEMDVAGAKKLEKLERG
KLRNLFRRK"
```

repeat_region	7563..7678	/rpt_family="CGG"n	Query Match	4.5%; Score 174.4; DB 8; Length 140681;
repeat_region	complement(10496..10524)	/rpt_family="AT_rich"	Best Local Similarity	49.9%; Pred. No. 1.1e-16;
repeat_region	10859..10900	/rpt_family="tandem"	Matches	696; Conservative 0; Mismatches 616; Indels 82; Gaps 7;
repeat_region	complement(11628..11727)	/rpt_family="Oryza"	Qy	1529
repeat_region	complement(11628..11713)	/rpt_family="Oryza"	Db	138438
repeat_region	complement(11631..11729)	/rpt_family="Oryza"	Qy	1589
repeat_region	complement(11631..11725)	/rpt_family="Oryza"	Db	138378
repeat_region	complement(11631..11725)	/rpt_family="TA"n	Qy	1649
repeat_region	11631..11711	/rpt_family="Oryza"	Db	138318
repeat_region	complement(11631..11709)	/rpt_family="Oryza"	Qy	1709
repeat_region	complement(11631..11707)	/rpt_family="Oryza"	Db	138258
repeat_region	complement(11631..11705)	/rpt_family="Oryza"	Qy	1766
repeat_region	complement(11631..11703)	/rpt_family="Oryza"	Db	138198
repeat_region	complement(11631..11701)	/rpt_family="Oryza"	Qy	1811
repeat_region	complement(11631..11697)	/rpt_family="Oryza"	Db	138138
repeat_region	complement(11631..11695)	/rpt_family="Oryza"	Qy	1871
repeat_region	complement(11631..11693)	/rpt_family="Oryza"	Db	138078
repeat_region	11631..11691	/rpt_family="Oryza"	Qy	1931
repeat_region	11631..11689	/rpt_family="Oryza"	Db	138018
repeat_region	11631..11687	/rpt_family="Oryza"	Qy	1991
repeat_region	11631..11685	/rpt_family="Oryza"	Db	137958
repeat_region	11631..11683	/rpt_family="Oryza"	Qy	2051
repeat_region	11631..11681	/rpt_family="Oryza"	Db	137898
repeat_region	11632..11732	/rpt_family="Oryza"	Qy	2111
repeat_region	complement(11632..11725)	/rpt_family="Oryza"	Db	137838
repeat_region	11634..11732	/rpt_family="Oryza"	Qy	2171
repeat_region	11642..11729	/rpt_family="Oryza"	Db	137783
repeat_region	complement(11644..11727)	/rpt_family="Oryza"	Qy	2231
repeat_region	complement(11646..11728)	/rpt_family="Oryza"	Db	137733
repeat_region	11648..11727	/rpt_family="Oryza"	Qy	2291

```
/rpt_family="Oryza"
complement(11648..11725)
/rpt_family="Oryza"
complement(11650..11727)
/rpt_family="Oryza"
11650..11725
/rpt_family="Oryza"
11652..11727
/rpt_family="Oryza"
complement(11652..11725)
/rpt_family="Oryza"
complement(11654..11727)
/rpt_family="Oryza"
11654..11725
/rpt_family="Oryza"
```

repeat_region	complement(11648..11725)	/rpt_family="Oryza"	Query Match	4.5%; Score 174.4; DB 8; Length 140681;
repeat_region	complement(11650..11727)	/rpt_family="Oryza"	Best Local Similarity	49.9%; Pred. No. 1.1e-16;
repeat_region	11650..11725	/rpt_family="Oryza"	Matches	696; Conservative 0; Mismatches 616; Indels 82; Gaps 7;
repeat_region	11652..11727	/rpt_family="Oryza"	Qy	1529
repeat_region	complement(11652..11725)	/rpt_family="Oryza"	Db	138438
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	1589
repeat_region	11654..11725	/rpt_family="Oryza"	Db	138378
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	1649
repeat_region	11654..11725	/rpt_family="Oryza"	Db	138318
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	1709
repeat_region	11654..11725	/rpt_family="Oryza"	Db	138258
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	1766
repeat_region	11654..11725	/rpt_family="Oryza"	Db	138198
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	1811
repeat_region	11654..11725	/rpt_family="Oryza"	Db	138138
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	1871
repeat_region	11654..11725	/rpt_family="Oryza"	Db	138078
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	1931
repeat_region	11654..11725	/rpt_family="Oryza"	Db	138018
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	1991
repeat_region	11654..11725	/rpt_family="Oryza"	Db	137958
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	2051
repeat_region	11654..11725	/rpt_family="Oryza"	Db	137898
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	2111
repeat_region	11654..11725	/rpt_family="Oryza"	Db	137838
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	2171
repeat_region	11654..11725	/rpt_family="Oryza"	Db	137783
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	2231
repeat_region	11654..11725	/rpt_family="Oryza"	Db	137733
repeat_region	complement(11654..11727)	/rpt_family="Oryza"	Qy	2291



/note="predicted by genemark hmm"

mRNA  
complement (<10339. .>11589)  
/locus tag="OSJNBb0018B10.10"  
complement (10339. .11589)  
/locus tag="OSJNBb0018B10.10"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAP54657.1"  
/db\_xref="GI:31433104"  
/translation="MAAEDWASLEQGLLHDFILRLDDANAARFVAVCSQWRAAAAG  
AGVVFPRPLAVAPORRHAWLPSSRRRRGRRRRFRKDDGVSRLRPRGGSPAPA  
AVLSVSRGWAMKHGESHVRAVSDEVPPLPCGSRSELSVPLPCCGSRVELRYE  
LSKVLDPDPLAAASGEWATAFMSHQQLDWHTRKRVAFCHAGDEWFLDIAIQA  
QRYGLFFQVSRVQVPRSCFPLLAGRRHRRRLFTSRVYKLEFADGSPVFGFTK  
VESIGELFVSRSHAFALPASGAASPAASRIASTTWSDYAARSASHRGRLAVSSP  
CHWTFISPGHSGFRIGFALAGO"  
complement (<13180. .>13660)  
/locus tag="OSJNBb0018B10.11"  
/note="predicted by genemark hmm"  
complement (join(<13180. .13414,13527. .>13660))  
/locus tag="OSJNBb0018B10.11"  
complement (join(13180. .13414,13527. .13660))  
/locus tag="OSJNBb0018B10.11"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAP54658.1"  
/db\_xref="GI:31433105"  
/translation="MAITMSVNRGRRAISGEGHLLGSBDTGKILWGGAPILS  
AGARICVFSALETSLAFLVPSRNPAPRFPVPAIAARVVRSTVAVALAVAVVN  
ADESPRRHRCCLLPTRR"  
<14243. .>22346  
/locus tag="OSJNBb0018B10.12"  
/note="predicted by genescan"  
join(<14243. .15321,16339. .16418,18176. .18376,19388. .19494,  
20446. .20632,20837. .20875,22300. .>22346)  
/locus tag="OSJNBb0018B10.12"  
join(14243. .15321,16339. .16418,18176. .18376,19388. .19494,  
20446. .20632,20837. .20875,22300. .>22346)  
/locus tag="OSJNBb0018B10.12"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAP54659.1"  
/db\_xref="GI:31433106"  
/translation="MAGGGSSLPADLLREVSGRILYSADHLHITHQVTHWRAATLPV  
SACRPWVAGALQRLVITIGDYSRLRPGGAERMDAGLRAPPGLRHCCGASRGWLA  
LTDARSPTRLALWELPGLABEIPCLSPITRSLSGDPLASPGWVAVSOLKRGEGQ  
KTLVCRGDAWMLFEHTGEIGSVFPHGKVYVIDIMSNIVLDCIDADGSAKCT  
QFFVTRVSMLCSDKFHPRGVRHLVACDGEVLLVLRMGHPSLAEIYKPEKGNH  
QLGGERVMDLGDHSLFVRGDTFSLSAKEFPAIKRNVYVADKNYKQOHWSIFHLE  
QVLLEELPYPEELKEDWNMPHSCKHCRQWISVTEIACLNHSHIYTHIDHGPGR  
FPVAILPWLAVAVRRPPRRPVSRRHPPPLFPLSLAAPGRATAPVVCRRJCAAT  
PVAGKGRDGAGRLRLSLPDSPRHGRGLPALRGVGAANSIEFPVLGRVAQDILTI  
PASTVSLSEVSTSGSLVSEFRSOUTPETTKALVCLQDWIRESGSSSELSTESIFDSK  
VVLQVSFEAQSP"  
<25677. .>28436  
/locus tag="OSJNBb0018B10.13"  
/note="similar tomato Cf 9 plant disease resistance  
polyprotein GI:7489045 (Lycopersicon esculentum)"  
join(<25677. .27248,27324. .>28436)  
/locus tag="OSJNBb0018B10.13"  
join(25677. .27248,27324. .28436)  
/locus tag="OSJNBb0018B10.13"  
/codon\_start=1  
/product="putative plant disease resistance polyprotein"  
/protein\_id="AAP54660.1"  
/db\_xref="GI:31433107"

gene  
/translation="MESKULEKLEGLIKMHHEIDWLSQGSLEKVTIKVLDLMDFV  
GVQVQCAELASMRKVCANCSQAGILSPGISQVSAGSATPPLNPPVPLVDGV  
DLGRFGDGIYRRHPLPMDPFRFDGSDVRIWLNMCETYFDMYQITQNFKYSAYVLH  
MSGNAQWYHSYKLVNNEVNSDQFRMAVATEFVGVRREKMSALDITLTQTFVTEYKQ  
QPDLYVQIIRVDPSPVGGKMLVTRFMNGLTEIRAAVVQLPDTVQSSAIAMQESV  
LASRATKALKGKFEVLPQYKSDVSSDTTPGQVDKGLDWAKQLKDYRRAQGLCFKCGD

gene  
/locus tag="OSJNBb0018B10.15"  
/note="similar tomato Cf 9 plant disease resistance  
polyprotein GI:7489045 (Lycopersicon esculentum)"  
join(<30126. .30676,30790. .31002,31382. .>33383)  
/locus tag="OSJNBb0018B10.15"  
join(30126. .30676,30790. .31002,31382. .33383)  
/locus tag="OSJNBb0018B10.15"  
/codon\_start=1  
/product="putative plant disease resistance polyprotein"  
/protein\_id="AAP54661.1"  
/db\_xref="GI:31433108"

mRNA  
/translation="MATARARVATGDIRRRRRKGVAKAAADGGGDRRRRVARVA  
TTHGESKKEKGEVLTGSDGRRRGATGEGHAAADGDDDLKNEKKKIREGDRHR  
RREPAPAGGGHGHPRHTGMALRRRTATKEGTWMTSATRTRRRRGGAAASLAAS  
GFSGNGGKRLHAGAMGSGDGGGGERGIRKSNPHLGRGRREGSGRGERGA  
RAAWATPFGGGDVGTAAVAGAARRRRRREVGSGDPDLGRTSQCDDTATQLAFDQLK

Query Match 4.5%; Score 174.4; DB 8; Length 30029;  
Best Local Similarity 49.9%; Pred. No. 1e-16;  
Matches 696; Conservative 0; Mismatches 616; Indels 82; Gaps 7;  
QY 1529 ACGTGCTCTCGGTGAGCGCATCTTCAACGCCGACGGCGGCTGTGAGGAGCAGAGGA 1588  
Db 295841 AGGACCTGTTTCGGGGAGCGGCATCTTCCCATCGACGGCGAGAGTGAAGCAGCAGGA 295782  
QY 1589 AGACGGCGAGTTTCAGTTCGCTCCCAAGAACCTGAGGGATTTTCAGCGCCATTTGTGTTCA 1648  
Db 295781 AGATCGCCAGCTAGACTTCTCCACAGGCGCCCTCCCGGACTTCAGCTGCGCGCTTCA 295722  
QY 1649 GAGAGTACTCCCTGAAGCTGTGCGGTATATCTGAGCCAGGCATCCCAAGCGAGGCAAGTTG 1708  
Db 295721 AGAGGAACGCCGCAAGCTGCGCGCATCGTCTCAACCCAGCGCGCTCAACCAATCCA 295662  
QY 1709 TGGACATGCACTGAGTCACTGCTC---CCTTCGCCATTCGCAACATGAGCATTTTCAACC 1765  
Db 295661 TGGACTTCCAGGTAAATTAAGCGATCGTAACACAAATGCTAATTAATGACGCGCGCG 295602  
QY 1766 TGAGACACGAGAGCTACTCT-----TGCCGATTTCAGGAATTTTACATGAGG 1810  
Db 295601 GTGAGACGGTGAGTGACGTACGGTTTGTGTCGAAGTTTGTTCAGGGTTTGTATGCTGAGA 295542  
QY 1811 ATGACGCTGGACTCACTGTCGAAGTTTGGGTTGGGGTTCGGGGTTCGAGATCGGCACTGTCGCCG 1870  
Db 295541 GCAACGATGGACTCCATCTTCCCATCGCATTCGGCGCAGACCTCAACACCTCGACGCGC 295482  
QY 1871 GATCTCCCGGAGACAGCTTCGCGCAGCGGTTTCGATGCGCGCAACATCATCGTCAAGCTG 1930  
Db 295481 TCCGCGGAGGGAGCGCGCTTCGCGCGCGGCTTCGACGACGCGCAGTGTACCATGCTC 295422  
QY 1931 CGGTTTCATCGACCGCGCTGTGGCGCATCAAGAGAGTTTCTTCCACGCTGGGTCAGAGCCCTC 1990  
Db 295421 CGTTACATCAGCCCGCTCTGGAAGCTGGCAAGGCTCCTCAACGCTGGCGTTCGAGGCCATG 295362  
QY 1991 CTAGCGCAGAGCATCAAGCTCGTGGACGAGTTTCACTACCGCTGATCCCGCGGGGAGGAG 2050  
Db 295361 CTCAGGAGAGAGTCAAGTCTGTCGACGAAATCGTGTACAGGCTCATCCGTCGCCAGTCC 295302  
QY 2051 GCCGAGATCGTCGAGCGCGCGCCAGCGGCAAAAGGAGAGGTACGTGCACTGACTGTT 2110  
Db 295301 GACGAGCTCTCAACTCACAGACTCCGTAAGCAACCCCTGAACTAAAGAGAGCTCCTAC 295242  
QY 2111 TTCGATTCTTCAGTTTCATCGTCTTGGCGCGGATGGACCTGATCTCTGATTGATTATATC 2170

KYTPHRCVAGGQIKAWOLEEVLTDILDAVIAAEQSDDEEDCHISINALTGAHLPTA  
ILLRALVSKVLLLLVDYSSHSFDYDNEAQMGLPLKPISTPLVANGDCLPCQVY  
VPGSWMIQHTFTYDMRVVITLGVSHDAILLGMNDLEQWEGMSCHWANKTLKFEQGSNI  
SLQGVQAAVPESEIAADDDVHVIPSSVQQLLTDVDFHDKPCLPKPHRVFEDHALFLVP  
DATPNCRPYRSPLOKDEIEKQKEMIDAGLITPSMSPYASPVLLVKKDGGSRFCV  
DYRLNALTIKSFPMPPVDELLDELSTGTFKFLDLKAGYHQIRMVETDEAKTAFKT  
HGFQFQFVMPFGLTNA PSTFOCLMNSVFAPFIRKFLVFMNDILVYSPDLSDSHHL  
QQVFLRQHQIFAKSKCTCTFACTOLEYLGXIIISDKGVSTDPEKTAAMLAWPVPTSVT  
ELMGFLGLTYRKFVQGSYIARPLILLKFLKKAFLWTDCHAQKFPNRSRRCIKIP  
VDPQGYTNDNDVDIQIHVLQTLKLYNAAEKKONRS"  
<30126. .>33383  
/locus tag="OSJNBb0018B10.15"  
/note="similar tomato Cf 9 plant disease resistance  
polyprotein GI:7489045 (Lycopersicon esculentum)"  
join(<30126. .30676,30790. .31002,31382. .>33383)  
/locus tag="OSJNBb0018B10.15"  
join(30126. .30676,30790. .31002,31382. .33383)  
/locus tag="OSJNBb0018B10.15"  
/codon\_start=1  
/product="putative plant disease resistance polyprotein"  
/protein\_id="AAP54661.1"  
/db\_xref="GI:31433108"

Db 295241 ACTCGGTTTGACT-----GACATATATATAGATGATCATCTTCGGATGTGTAACTAAG 295187  
QY 2171 CGTGTGACTTGTGAGGACAAATTAATCGGAGATGAAGACACGACATCTCTGTCACGGTT 2230  
Db 295186 C-----TGAATTTGTATCTGTACAGGTTTCGAGGAGGATATCTATCAAGATT 295137  
QY 2231 CATCGAGCTAGGCGGCGGCGGACGACGCGCGCGGCTTCGGGGACGACAAAGAGCTCGG 2290  
Db 295136 TCTCCAGGCAACGA-----CCAGCATCTCTGTGTCGATTACAAAGTACCTCGG 295089  
QY 2291 GGAAGTGTGCTCAACTTCTGTATCGCGGCGGAGACGACGCGGACGACGCTGTCTG 2350  
Db 295088 AGACATCACTACCTGAACTTGTATGTCATAGCCGCAAGGACACACAGCTGGGCGCTTGT 295029  
QY 2351 GTTCAGGCATGGCCATGTCCTCCACCGGACGCTGGCGGAGAGCTCGCGGCGAGCTGTG 2410  
Db 295028 GTTCTGTATAGTGTGCAAGCACCGGAGCTCCAGGAGAGATCTGCCACGAGCCAT 294969  
QY 2411 CGGTTTCGAGGCGGCGGCGGCGGAGGAGGCGTCTGCGCTGCTGCTCCCTGCGGCGGCG 2470  
Db 294968 GGTGGCCACAGCGCGCGGACACCGCTTCCGTCGACGAGTCTTTCGACAGCTTGA 294909  
QY 2471 TGACGCGGACGACGAGCGTTCGCGCGCGGCGGCGAGTTCGCGGCTCTCACCTA 2530  
Db 294908 CNAAGCGCTGAACAAATGACATACCTGACGCGGCGACCTGACGAGACGCTCGGCTTTA 294849  
QY 2531 CGACAGCTCGGCAAGCTGTCTACCTCCACGCGCTGCGTCACGAGACGCTCGGCTGTA 2590  
Db 294848 CCCTTCAGTTCCTCAATGTTGATTAATATGATGATCAGTTTCACTA----- 294805  
QY 2591 CCGCGCGCTCCCTCAGTGAGCGCGCGGCGGACGCGACCTCCGTTCCAGACGACGATG 2650  
Db 294804 -----CCATCGAGCTAAATTAATCAACGACATTTCTCTCATGAGTGGGTAAACAA 294754  
QY 2651 CAGTGAGTGACCTGAATGCAATGCATGCACTTCGCGCGGCGGAGGACCCCAAGGGA 2710  
Db 294753 CGTCTCTGTGTAG-----ATGATCTTTTCTTTTCAGGAAACGACGAT 294706  
QY 2711 TCCTGAGGACGACGCTGCTCGCGGAGGAGGAGGAGGCGGCGGCGGATGGTGACGT 2770  
Db 294705 GCTTTTCAGACGATGATTGCCCCAAGCTTTTAACTGACGAGGAGGACATCTGTTCT 294646  
QY 2771 AGTGCCCTACTCGATGGGCGGATGGAGTAACTTGGGCGGCGGCGGAGGCTTCC 2830  
Db 294645 TCATCCCTACGCGATGGGCGGATGGAGAGCTTGTGGGCAAGACGCTGAATCTTC 294586  
QY 2831 GGCGGAGCGGTGGATCAACGAGGATGGCGCTTCGCAACGCGTTCGCTTCAAGTTCA 2890  
Db 294585 GGCCTGAAAGTTGGCTCGACGAAACGCGCTCTTCAGCGGAAAGCGGCTTCAAGTTTA 294526  
QY 2891 CGCGGTTCCAGGCG 2904  
Db 294525 CAGCAATTCAGGTG 294512

## RESULT 15

CNS08C9V

LOCUS

DEFINITION

Oryza sativa chromosome 12, . BAC OSJNBa0041M09 of library OSUNBa  
from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza

ACCESSION

AL772417

VERSION

AL772417.3

KEYWORDS

HTG.

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzoideae; Oryza.

REFERENCE

1 (bases 1 to 146519)

AUTHORS

Choisne, N., Orjeda, G., Cattolico, L., Demange, N., Wincker, P., Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,

Weissenbach, J. and Quetier, F.  
Oryza sativa chromosome 12 sequencing  
Unpublished  
2 (bases 1 to 146519)  
Genoscope.  
Direct Submission  
Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
On Oct 2, 2002 this sequence version replaced gi:22003295.  
Center: Genoscope / Centre National de Sequencage  
Center code: GS  
Web site: http://www.genoscope.cns.fr/  
Contact: Seqref@genoscope.cns.fr

-----  
The following sequence is oriented from the T7 to the SP6 end.  
Upstream BAC (overlapping the T7 end) : OSJNBa0037B01 (AC=AL928751)  
Downstream BAC (overlapping the SP6 end) : OSJNBa0068D20  
(AC=AL731873) ----- Finishing boundaries  
FINISHED SEGMENT STARTS AT BASE 1  
FINISHED SEGMENT ENDS AT BASE 146519  
-----

## FEATURES

source

1. 146519  
Location/Qualifiers  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:39947"  
/chromosome="12"  
/clone="OSJNBa0041M09"  
/clone\_lib="OSUNBa"

## ORIGIN

Query Match 4.2%; Score 162.4; DB 8; Length 146519;  
Best Local Similarity 56.5%; Pred. No. 6.9e-15;  
Matches 383; Conservative 0; Mismatches 286; Indels 9; Gaps 4;  
QY 2399 CCGGAGCTGTGCGGTTGAGGCGGAGCGCGCGCGGAGGCGGCTCGGCTGTGCC 2458  
Db 8689 CTTGTTGTCGCGATGCGCGCGCGGCTACCCGACGAGGCGATCTGCGACATGTGT 8748  
QY 2459 CTGCGCGCGCTGAGCGCGACGACGAGGCTTCGCGCGCGCGGCGGCTGCGGG 2518  
Db 8749 GAGCTTCATCATGCGCGCGCGGACGACGCTGCTCGCGGCGC---TGACGTGTTCTTCTG 8805  
QY 2519 CTTCTCTACCTACGACGCTCGGCAAGCTGCTACCTTCCACGCGCTGCTCACGAGAC 2578  
Db 8806 GCTCATGACGCGCCACCGCGAGCTGGAGCGGAGGTTCTTGGACGAGTTCGACGCTGCAT 8865  
QY 2579 GCTCGGCTGTACCGCGCGCTCCCTCAGTGAGGCGCGCGGACACGCGACCTCCGTTCA 2638  
Db 8866 GGGCGACGCGCGGATGCGGCTGCGCTCGCGCGCTGGAC--CTCAGGCGCTCGCGCAG 8923  
QY 2639 GAGCACGATGACGTGAGTGACCTGATGCAATGCAATGCAATGCAATGCAATGCAATG 2697  
Db 8924 GCGCGGTGCTCCACGCGCGCTCTCGGAGACGATGCGGCTGTACCCGCGGTGGGCTGG 8983  
QY 2698 GACCCCAAGGGATCTCGGAGGACGACGCTGCTCGCGGACGAGGAGGAGGCGCGG 2757  
Db 8984 GACTCGAACACGCGCGGAGGACGACGCTGCTCCGCGGACGCGCGCTGGGCGCGGCG 9043  
QY 2758 GGGATGTTGACGTGCTGCTTCTGATGGGCGGATGAGTACACTTGGGCGCGCGGAC 2817  
Db 9044 GACCGGCTCACCTACTTCCCTTACGGGATGGGAGATGGAGGCGCATCTTGGGCGCGG 9103  
QY 2818 GCGGCGAGCTTCGCGCGGAGCGGTGGATCAACGAGATGGCGCGTTCGCGCAACGCTCG 2877  
Db 9104 GCGGCGAGTTGAGGCGGAGGAGTGGCTTGGCGCGCGCGCGCGCGCGCGCGGCTGTCG 9163  
QY 2878 CGGTTCAAGTTACGCGGCTTCCAGGCGGCGCGGAGGATCTGCTGGGCAAGGACTCGGCG 2937  
Db 9164 CCGTTCAAGTACCGGTTTCCAGGCGGCGCGGACGCTGCTCGGCAAGGACATGGCG 9223

QY	2938	TACCTGCAGATGAAG---	ATGGCGCTGGCCATCCTTTGCGCTTCTACAGCTTCCGGCTG	2994
Db	9224	TTCTGTGAGATGAAGTTCTGTGGCCAGCGCCGTCTCCGGCGGTTGAGCTCCGCCCGGTG	9283	
QY	2995	CTGGAGGGGCAACCGGTGAGTACCGCATGATGACCATCTCTCTCCATGCGCGCACGGGCTC	3054	
Db	9284	GCGCCGAGGGGCTCGCCGAGTTCTGTGCGGCTGTCTACCGCGCACATGGCCGGGGGCTC	9343	
QY	3055	AAGTCCGGCTCTTAGG	3072	
Db	9344	AAGGTGACGGTGAGGAGG	9361	

Search completed: March 7, 2004, 15:19:58  
Job time : 9799 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2004, 22:02:35 ; Search time 5017 Seconds  
(without alignments)

11344.891 Million cell updates/sec

Title: US-10-021-657-1

Perfect score: 1906

Sequence: 1 gaattcgacaggggaagc.....aaaaaaaaaactcgag 1906

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_nam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gsl:\*

29: gb\_gsl2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	818.4	42.9	997	11	AY110120	Zea mays
C 2	651.2	34.2	1002	29	CG344416	CG344416 OGLDH18TV
C 3	619.6	32.5	887	29	CG611162	CG611162 OGLDH18TV
C 4	579	30.4	833	29	CG317324	CG317324 OG0CS18TH

5	573.4	30.1	720	13	BQ840694	BQ840694 WHE4202_E
6	569	29.9	859	29	CG317335	CG317335 OG0CS18TV
7	554	29.1	563	10	AW519943	AW519943 660053C10
8	540	28.3	880	29	CG047034	CG047034 PUFWN66TB
9	533.8	28.0	578	12	BI075441	BI075441 IPI_20_E1
10	527.4	27.7	879	29	CG685050	CG685050 OGUFB46TV
11	509.4	26.7	585	12	BI075273	BI075273 IPI_18_C1
12	482	25.3	606	10	BE494080	BE494080 WHE1277_F
13	473	24.8	748	28	BZ532472	BZ532472 OGA1V02TC
14	463.6	24.3	517	10	BQ841568	BQ841568 WHE4212_E
15	461.4	24.2	591	13	AW424821	AW424821 660039C10
16	457.4	24.0	681	29	CG371414	CG371414 OGXBV64TH
17	454.2	23.8	819	28	CC368220	CC368220 PUHOP96TB
18	454.2	23.8	882	29	CC635363	CC635363 OGKAW86TV
19	454.2	23.8	991	29	CG047035	CG047035 PUFWN66TD
20	448.2	23.5	518	12	BG948686	BG948686 IPI_12_G1
21	445.6	23.4	620	13	BU996440	BU996440 HMI3J19r
22	441.6	23.2	625	13	BQ840748	BQ840748 WHE4203_A
23	438.2	23.0	864	29	CG344408	CG344408 OGDH18TH
24	427.4	22.4	773	28	CC368222	CC368222 PUHOP96TD
25	427.4	22.4	887	29	CG067339	CG067339 PUICF65TD
26	420.4	22.1	458	12	BI246000	BI246000 IPI_66_D0
27	417.4	21.9	509	12	BI099541	BI099541 IPI_43_E0
28	409.6	21.5	470	12	BI246162	BI246162 IPI_67_G0
29	406.6	21.3	615	29	CG611171	CG611171 OGLEJ88TV
30	394.4	20.7	649	12	BG948366	BG948366 IPI_18_G1
31	391.4	20.5	514	10	BF292135	BF292135 WHE2207_B
32	388.2	20.4	408	12	BM500648	BM500648 PAC000000
33	368.8	19.3	469	12	BG274193	BG274193 WHE2230_F
34	361	18.9	535	10	BF292000	BF292000 WHE2203_D
35	344.2	18.1	406	12	BI141660	BI141660 IPI_55_D0
36	338.2	17.7	400	12	BI141132	BI141132 IPI_43_E0
37	330.2	17.3	848	29	CG238851	CG238851 OGWER88TV
38	330	17.3	712	29	CG067338	CG067338 PUICF65TB
39	324.4	17.0	679	14	CD491751	CD491751 WHE2207_B
40	308.2	16.2	383	12	BG274260	BG274260 WHE2226_D
41	285	15.0	645	9	AU239680	AU239680 AU239680
42	284.2	14.9	794	29	CC685043	CC685043 OGUFB46TH
43	274	14.4	534	29	CG273559	CG273559 OGWHN30TH
44	272	14.3	526	12	BI074975	BI074975 IPI_18_C1
45	271	14.2	477	12	BI245608	BI245608 IPI_67_G0

## ALIGNMENTS

RESULT 1	AY110120/c	AY110120	937 bp	mRNA	linear	HTC 17-OCT-2002
LOCUS	Zea mays	CL58113.1	937 bp	mRNA	linear	HTC 17-OCT-2002
DEFINITION	Zea mays	CL58113.1	937 bp	mRNA	linear	HTC 17-OCT-2002
ACCESSION	AY110120	AY110120	937 bp	mRNA	linear	HTC 17-OCT-2002
VERSION	AY110120.1	GI:21214224	937 bp	mRNA	linear	HTC 17-OCT-2002
KEYWORDS	HTC	HTC	937 bp	mRNA	linear	HTC 17-OCT-2002
SOURCE	Zea mays	Zea mays	937 bp	mRNA	linear	HTC 17-OCT-2002
ORGANISM	Zea mays	Zea mays	937 bp	mRNA	linear	HTC 17-OCT-2002
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.					
AUTHORS	Hayney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.					
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Zea. Design of Overgo Probes					
JOURNAL	Unpublished (2002)					
REFERENCE	2 (bases 1 to 997)					
AUTHORS	Coe,E.H.					
TITLE	Direct Submission					
JOURNAL	Submitted (25-Apr-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA					
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, the www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the					

maize cDNA sequences is either Virginia walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES source Location/Qualifiers 1..997 /organism="Zea mays" /mol\_type="mRNA" /db\_xref="MaizeDB:632508" /db\_xref="taxon:4577" /clone\_lib="Maize Mapping Project/DuPont Consensus Library" /note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

ORIGIN Query Match 42.9%; Score 818.4; DB 11; Length 997; Best Local Similarity 86.2%; Pred. No. 1.7e-96; Matches 880; Conservative 0; Mismatches 116; Indels 25; Gaps 3; QY 845 AACAGGAGAGATGAAGACGACATCTGTCTGTCAGGTTCATCGAGCTGGCGGAGCGCGG 904 Db 997 AAAAGGGGAGATGAAGACGACATCTGTCTGTCAGGTTCATCGAGCTGGCGGAGCGCGG 938 QY 905 ACGACGGCGCGCTTCGGGGAGGATAGAGCTCCGGGAGCTGGTCAACTTCGTA 964 Db 937 NNN 878 QY 965 TCGCGGGGCGGACGACGCGGACGACGCTGTCTGTCTGTTTACGACATGGCCATGTCCC 1024 Db 877 TCGCGGGGCGGACGACGCGGACGACGCTGTCTGTCTGTTTACGACATGGCCATGTCCC 818 QY 1025 ACCGGACGTGGCGGAGAGCTGGCGGAGCTGGTGGAGCTGGTGGAGCTGGTGGAGCTGG 1084 Db 817 ACCGGACGTGGCGGAGAGCTGGCGGAGCTGGTGGAGCTGGTGGAGCTGGTGGAGCTGG 758 QY 1085 GCGAGGAGGCTCAGCTGTCTGTCTGGGCGGCTGACGCGGACGACAGGCTGG 1144 Db 757 NNN 698 QY 1145 CCGCGCGCTGGCGAGTTCGGGGGCTCTCCTACCTACGACGCTCGCAAGCTGGTCT 1204 Db 697 CCGCGCGCTGGCGAGTTCGGGGGCTCTCCTACCTACGACGCTCGCAAGCTGGTCT 638 QY 1205 ACCTCAGCGCTGGCTCAGGAGAGCTGGCTGTACCCCGCTGCTGCTGCTGCTGCTGCTG 1264 Db 637 ACCTCAGCGCTGGCTCAGGAGAGCTGGCTGTACCCCGCTGCTGCTGCTGCTGCTGCT 578 QY 1265 AGGGGATCTGGAGGACGAGTGTCTCCGAGCGGAGAGGAGTGGAGGCGCGGGATGG 1324 Db 577 AGGGGATCTGGAGGACGAGTGTCTCCGAGCGGAGAGGAGTGGAGGCGCGGGATGG 518 QY 1325 TGACGTACGTGGCTTACTCGATGGGCGGATGAGTACAACTGGGCGCCCGGAGCGGCGGA 1384 Db 517 TGACGTACGTGGCTTACTCGATGGGCGGATGAGTACAACTGGGCGCCCGGAGCGGCGGA 458 QY 1385 GCTTCGGCGGAGCGGTGGATCAAGAGATGGCGGTTCGCAACGCGTTCGCAACGCGTTC 1444 Db 457 GCTTCGGCGGAGCGGTGGATCAAGAGATGGCGGTTCGCAACGCGTTCGCAACGCGTTC 398 QY 1445 AGTTACGCGGCTTCAGGGGCGGCGGAGGATCTGCTGGGCAAGGACTCGGCGTACTGTC 1504 Db 397 AGTTACGCGGCTTCAGGGGCGGCGGAGGATCTGCTGGGCAAGGACTCGGCGTACTGTC 338 QY 1505 AGATGAAGATGGCGCTGGCGCATCTCTTGGCTTCTACAGTTCGGCTGCTGGAGGGCC 1564 Db 337 NNN 278 QY 1565 ACCGGTGCAGTACCGCATGATACCATCTCTCCATGGGCGACGCGCTCAAGTCCGCG 1624

Db 277 ACCCGTGCAGTACCGCATGATACCATCTCTCCATGGGCGACGCGCTCAAGTCCGCG 218 QY 1625 TCTCTAGGGCGCTGTGATGTGATGGCGATTTGGATATGATATCGTCCCGCTTAATCCAC 1684 Db 217 TCTCTAGGGCGCTGTGATGTGATGGCGATTTG-----GGATATCATCCGCTTAATCC-- 165 QY 1685 GACAAATAACGCTCGTGTTCATAAATTGTCATGCAATGCAATGCAATGCAATGCAATG 1744 Db 164 -----TNN 122 QY 1745 CATTTGGTGGCTTAAAGCTTAAAGAACTCGTCCGCTTTCGGAACCAACCAATCATCAC 1804 Db 121 CATTTGGTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAA 62 QY 1805 TAGTGTCTTTTGTACTTACTTCTCCTCAGTGGAAAGTGTAGTGACAGCAT-ACAAGTTCAT 1863 Db 61 TAGTGTCTTTTGTACTTACTTCTCCTCAGTGGAAAGTGTAGTGACAGCATACAAGTTCAT 2 QY 1864 T 1864 1 T 1

RESULT 2 CG344416/c LOCUS CG344416 1002 bp DNA linear GSS 26-AUG-2003 DEFINITION OG1D18TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0736C12, genomic survey sequence.

ACCESSION CG344416 VERSION CG344416.1 GI:34261682 KEYWORDS GSS. SOURCE Zea mays ORGANISM Zea mays Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 1002)

REFERENCE 1 White, C.A., Quackenbush, J., Van Aken, S., Uterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nurnberg, A., Robbins, D. and Lakey, N. Consortium for Maize Genomics Unpublished (2002) Other\_GSSs: OGI018TH Contact: Cathy Whitelaw TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-838-5843 Fax: 301-838-0208 Email: whitelaw@tigr.org Seq primer: 1F Class: sheared ends.

FEATURES source Location/Qualifiers 1..1002 /organism="Zea mays" /mol\_type="genomic DNA" /strain="B73" /db\_xref="taxon:4577" /clone="ZMMBMA0736C12" /clone\_lib="ZM 0.7 1.5 KB" /note="Vector: pBCSK-; site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"

ORIGIN Query Match 34.2%; Score 651.2; DB 29; Length 1002; Best Local Similarity 83.6%; Pred. No. 7.2e-75; Matches 833; Conservative 0; Mismatches 3; Indels 160; Gaps 2; QY 9 CACGAGGAGAGTCACTACCTACGCGCGGACGCGCATTCCTCCACATGAGGAGGCC 68 Db 996 CATGAGGAGAGTCACTACCTACGCGCGGACGCGCATTCCTCCACATGAGGAGGCC 937 QY 69 TCACAGTACATCGCGCTCTCTCTGTTGTCCTCTCATGATCCTGGTCCAGAGGTGGAG 128

Db 936 TCACAAAGTACATCGCGCTCCTTCTGGTGTGCTCTCATGGATCCTGCTCCAGAGGTGGAG 877  
 Qy 129 CTTGAGGAAGCAAGAGCCCGAGATCATGGCCAGTTCATCGGCGCAACGGTGGAGAGCT 188  
 Db 876 CTTGAGGAAGCAAGAGCCCGAGATCATGGCCAGTTCATCGGCGCAACGGTGGAGAGCT 817  
 Qy 189 GAGGAATACACCGGATGACAGCTGGTGTGCGGTACCTGTACGGCACAGGACAGT 248  
 Db 816 GAGGAATACACCGGATGACAGCTGGTGTGCGGTACCTGTACGGCACAGGACAGT 757  
 Qy 249 GACGTCGACATGCCGTTCTCTTCTACACTACATCGCTGACCGGTGAATGTCGAGCA 308  
 Db 756 GACGTCGACATGCCGTTCTCTTCTACACTACATCGCTGACCGGTGAATGTCGAGCA 697  
 Qy 309 TGTCTCAAGACTAACTTCCACCAATACCCCA-----AGGAA 346  
 Db 636 CAGTCTTCGGAATCAGAGCTGAAGCTGAATCGAATGTGCTGAACACCGTGTAGGAA 577  
 Qy 347 TCGTGTACAGATCCTATCATGAGCTGCTCCTCGTGACGGATCTTCAACGCGACGCG 406  
 Db 576 TCGTGTACAGATCCTATCATGAGCTGCTCCTCGTGACGGATCTTCAACGCGACGCG 517  
 Qy 407 AGCTGTGAGGAAGCAGAGGAGAGCGGAGTTCGAGTTTCCCTCCAAAGAACCTGAGG 466  
 Db 516 AGCTGTGAGGAAGCAGAGGAGAGCGGAGTTCGAGTTTCCCTCCAAAGAACCTGAGG 457  
 Qy 467 ATTTGAGCGCCATTGTGTTACAGAGTACTCCCTGAAGCTGTGGGTATGATGAGCCAGG 526  
 Db 456 ATTTGAGCGCCATTGTGTTACAGAGTACTCCCTGAAGCTGTGGGTATGATGAGCCAGG 397  
 Qy 527 CATCAAGGCGAGCAAGATTTGTCGACATG----- 555  
 Db 396 CATCAAGGCGAGCAAGATTTGTCGACATGCAAGGTGAGATATCATGTTCCCTTCCCATG 337  
 Qy 556 -----CAGGAATTTTACA 568  
 Db 336 CCAACATGAGCATTTCAACCTGAGACAGAGAGTACCTTCCGATTCAGGAATTTTACA 277  
 Qy 569 TGAGGATGAGCTGACATCTCATCTGCAAGGTTGGTTCGGGGTGGAGATCGGACGCTGT 628  
 Db 276 TGAGGATGAGCTGACATCTCATCTGCAAGGTTGGTTCGGGGTGGAGATCGGACGCTGT 217  
 Qy 629 CGCAGATCTCCCGAGAACAGCTTCGCGCAGGCGTTCGATGCGCGCAACATCATCA 688  
 Db 216 CGCCAGATCTCCCGAGAACAGCTTCGCGCAGGCGTTCGATGCGCGCAACATCATCA 157  
 Qy 689 CGCTCGGTTTCATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTGAGG 748  
 Db 156 CGCTCGGTTTCATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTGAGG 97  
 Qy 749 CCCTCCTAGCGAGAGCATCAAGCTCTGTGGAGAGTTCACCTACAGCTGATCCGCGCGA 808  
 Db 96 CCCTCCTAGCGAGAGCATCAAGCTCTGTGGAGAGTTCACCTACAGCTGATCCGCGCGA 37  
 Qy 809 GGAAGCCGAGATCTCGAGGTCGCGGCGCAGCGCA 844  
 Db 36 GGAAGCCGAGATCTCGAGGTCGCGGCGCAGCGCA 1

RESULT 3  
 LOCUS CC611162/c  
 DEFINITION OGLB88TH ZM\_0.7\_1.5\_KB zea mays genomic clone ZM8MA0315P08,  
 genomic survey sequence.  
 ACCESSION CC611162  
 VERSION CC611162.1 GI:31972583  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 887)  
 Whitehead, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other GSSs: OGLB88TV  
 Contact: Cathy Whitelaw

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 1..887  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZM8MA0315P08"  
 /clone\_lib="ZM\_0.7\_1.5\_KB"  
 /notes="Vector: pBSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

LOCATION/Qualifiers  
 1..887  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZM8MA0315P08"  
 /clone\_lib="ZM\_0.7\_1.5\_KB"  
 /notes="Vector: pBSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

## ORIGIN

Query Match 32.5%; Score 619.6; DB 29; Length 887;  
 Best Local Similarity 97.1%; Pred. No. 9e-71;  
 Matches 631; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
 Qy 1232 TCGCGCTGTACCGCGCTCCTCAGACCCCAAGGGGATCTCTGGAGACGACGTGTGTC 1291  
 Db 851 TGCATATGCACTTTCGCGCGCGCAGGACCCCAAGGGGATCTCTGGAGACGACGTGTGTC 792  
 Qy 1292 CGGACGGAGCAAGGTGAGGCGCGCGGATGTGACGTAGTGCCTTCTCGATGGGC 1351  
 Db 791 CGGACGGAGCAAGGTGAGGCGCGCGGATGTGACGTAGTGCCTTCTCGATGGGC 732  
 Qy 1352 GGATGGAGTACAATGGGCGCGCGAGCTTCGCGCGGAGCGGCGGATCAACG 1411  
 Db 731 GGATGGAGTACAATGGGCGCGCGAGCTTCGCGCGGAGCGGCGGATCAACG 672  
 Qy 1412 AGGATGGCGGTTCCGCAACGCGTTCGCGGTTCAAGTTCA CGGCGTTCAGCGCGGCGCA 1471  
 Db 671 AGGATGGCGGTTCCGCAACGCGTTCGCGGTTCAAGTTCA CGGCGTTCAGCGCGGCGCA 612  
 Qy 1472 GGATTCGCTGGGCAAGGACTCGCGGTACCTGCGAGATGAAGATGGCGTGGCCATCTCT 1531  
 Db 611 GGATTCGCTGGGCAAGGACTCGCGGTACCTGCGAGATGAAGATGGCGTGGCCATCTCT 552  
 Qy 1532 TCCGCTTCTACAGCTTCGCGGTCTGGAGGGGCAACCGGTGCAGTACCGCATGATGACCA 1591  
 Db 551 TCCGCTTCTACAGCTTCGCGGTCTGGAGGGGCAACCGGTGCAGTACCGCATGATGACCA 492  
 Qy 1592 TCCTCTCCATGGCGACGGCTCAAGGTTCGCGTCTTAGGGCGGTCTGATGTCATGCGG 1651  
 Db 491 TCCTCTCCATGGCGACGGCTCAAGGTTCGCGTCTTAGGGCGGTCTGATGTCATGCGG 432  
 Qy 1652 ATTTGGATATGGATATGTCGCGCTTAATCCACGACAAATACCGCTCGTGTACAAATTT 1711  
 Db 431 ATTTGGATATGGATATGTCGCGCTTAATCCACGACAAATACCGCTCGTGTACAAATTT 372  
 Qy 1712 GCATGCATGCATGTAAGGAAAGCGATGGGTTTCAATGGTGGCTTGGCTTAAGCCTTAAA 1771  
 Db 371 GCATGCATGCATGTAAGGAAAGCGATGGGTTTCAATGGTGGCTTGGCTTAAGCCTTAAA 312  
 Qy 1772 AACTCCGTGGGTCTTGGCAACCAACCATCATCACTAGTGTGTTGTTACTTACTCTCAGTG 1831  
 Db 311 AACTCCGTGGGTCTTGGCAACCAACCATCATCACTAGTGTGTTGTTACTTACTCTCAGTG 252



```

/dev_stage="Pre-meiotic anthers"
/lab_host="E. coli SOLR"
/clone_lib="Aegilops segetoides pre-meiotic anther cDNA
library"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in a growth
chamber at the University of California, Davis (Akhunov in
Devorak's lab). Pre-meiotic anthers were harvested, total
RNA and poly(A) RNA were prepared, a cDNA library was
made, and the cDNA clones were in vivo excised to give
pBluescript phagemids at the University of California,
Davis (Akhunov in Devorak's lab). Plasmid DNA preparations
and cDNA sequencing were performed in the OD Anderson lab
(all other authors)."

```

ORIGIN

	Query Match	30.1%;	Score 573.4;	DB 13;	Length 720;
	Best Local Similarity	88.6%;	Pred. No. 8.8e-65;		
	Matches 635;	Conservative 0;	Mismatches 76;	Indels 6;	Gaps 1;
Qy	276	CACCTACATCGCTGACC CGGTGAATGTCGAGCATGTCTCAAGACTAACTTCAACAATT A	335		
Dd	4	CACGAGATCGCCGACCGGGTTAAGTCGAGCATGTGCTCNAGACCACATTCACAAATT A	63		
Qy	336	CCCCAAGGGAAATCSTGTACAGATCCTACATGGAAGTGCTCTCGGTGACGCATCTCTCAA	395		
Dd	64	CCCCAAGGGGAGGTGTACAGTCTCTACATGGAAGTGCTCTCGGCACGCATATTCAA	123		
Qy	396	CGCCGACGGGAGCTGTGGAGGAACGAGAGAAGACGGCGAGTTTCGAGTTCGCCCTCAA	455		
Dd	124	CGCCGACGGCGAGCTCTGGAGGAACGAGAGAAGACGGCGAGCTTCGAGTTCGCTTCAA	183		
Qy	456	GAACTTGAGGATTTTCAGCGCATTTGTTTCAGAGAGTACTCCCTGAAGCTGTCCGGTAT	515		
Dd	184	GAACTTGAGACATTTACAGCAGATCTGTTTCAGGGAGTACTCCCTGAAGCTGTCCAGCAT	243		
Qy	516	ACTGAGCCAGGCATCCAAGGCAGGCAAAAGTTGTGGACATGCAAGAACCTTTACATGAGAT	575		
Dd	244	CCTGAGCCAGSCTTCAAGGCGGCAAGTTGTGGACATGCAAGAGCTGTACATGAGGAT	303		
Qy	576	GACCTGGAATTCATCTGTCAAGATTGGTTCCGGGTTCAGAGTCGACAGCTGTGCCAGA	635		
Dd	304	GACCTGGAATTCGATCTGTCAAGTTGGGTTCCGGGTTCGAGATCGCACGCTGTCCCGGA	363		
Qy	636	TCTCCCAGAAACAGCTTCGCGCAGGCGTTCTGATGCGCCCAACATCATCATCAGCTGCG	695		
Dd	364	GCTGCGGAGAACAGCTTCGCGCAGGCGCTTCGACGCGCCAAATCATCTGTGACGCTGCG	423		
Qy	696	GTTTCATCGACCCGCTGTGGCGCATCAAGAGTTCTTCCACGTCCGGTTCAGAGGCCCTCT	755		
Dd	424	GTTTCATCGACCCGCTGTGGCGGTGAAGAAGTTCTCTGCACCTCGCTCGGAGGCGCTGCT	483		
Qy	756	AGCGGAGAGATCAAGCTCTGTGGAGAGTTCACTTACAGCGTGAATCCGCGGAGGAAGGC	815		
Dd	484	GGAGCAGAGCATCAAGCTCTGTGAGAGTTCACTTACAGCGTTCATCCGCGGCGCAAGGC	543		
Qy	816	CGAGATCGTTCGAGTTCGCGGCCAGCGGCAACAGGAGAAAGATGAACACGACATCTCTGTC	875		
Dd	544	CGAGATCGTTCGAGGCCCGGCCAGCGGCAAGCAGGAGAAAGTTCAGACAGACGTAATCTGTC	603		
Qy	876	ACGGTTTCATCGAGCTGGGCGAGGCCGCGACGACGCGCGCGC-----TTTCGGGAGCA	929		
Dd	604	GCGGTTTCATCGAGCTGGGCGAGGCCGCGGCGACGACGCGCGCAGCCTGTTTCGGGAGCA	663		
Qy	930	TAAGAGCTCCGGGACGTGTGCTCAATCTTGATTCGCGGGCGGGACACGACGGC	986		
Dd	664	CAAGGCGCTCCGCGACGTGGTCTCAATCTTGATTCGCGGGCGGGACACACGCGC	720		

RESULT 6	CG317335	LOCUS	CG317335	859 bp	DNA	linear	GSS 26-AUG-2003
CG317335							
DEFINITION	OG0CG318TV	ZM	0.7	1.5 KB	Zea mays genomic clone ZMMBMa0698D11,		

ACCESSION VERSION KEYWORDS SOURCE ORGANISM	REFERENCE AUTHORS	TITLE JOURNAL COMMENT
<p>1</p> <p>2</p> <p>3</p> <p>4</p> <p>5</p> <p>6</p> <p>7</p> <p>8</p> <p>9</p> <p>10</p> <p>11</p> <p>12</p> <p>13</p> <p>14</p> <p>15</p> <p>16</p> <p>17</p> <p>18</p> <p>19</p> <p>20</p> <p>21</p> <p>22</p> <p>23</p> <p>24</p> <p>25</p> <p>26</p> <p>27</p> <p>28</p> <p>29</p> <p>30</p> <p>31</p> <p>32</p> <p>33</p> <p>34</p> <p>35</p> <p>36</p> <p>37</p> <p>38</p> <p>39</p> <p>40</p> <p>41</p> <p>42</p> <p>43</p> <p>44</p> <p>45</p> <p>46</p> <p>47</p> <p>48</p> <p>49</p> <p>50</p> <p>51</p> <p>52</p> <p>53</p> <p>54</p> <p>55</p> <p>56</p> <p>57</p> <p>58</p> <p>59</p> <p>60</p> <p>61</p> <p>62</p> <p>63</p> <p>64</p> <p>65</p> <p>66</p> <p>67</p> <p>68</p> <p>69</p> <p>70</p> <p>71</p> <p>72</p> <p>73</p> <p>74</p> <p>75</p> <p>76</p> <p>77</p> <p>78</p> <p>79</p> <p>80</p> <p>81</p> <p>82</p> <p>83</p> <p>84</p> <p>85</p> <p>86</p> <p>87</p> <p>88</p> <p>89</p> <p>90</p> <p>91</p> <p>92</p> <p>93</p> <p>94</p> <p>95</p> <p>96</p> <p>97</p> <p>98</p> <p>99</p> <p>100</p>	<p>1</p> <p>2</p> <p>3</p> <p>4</p> <p>5</p> <p>6</p> <p>7</p> <p>8</p> <p>9</p> <p>10</p> <p>11</p> <p>12</p> <p>13</p> <p>14</p> <p>15</p> <p>16</p> <p>17</p> <p>18</p> <p>19</p> <p>20</p> <p>21</p> <p>22</p> <p>23</p> <p>24</p> <p>25</p> <p>26</p> <p>27</p> <p>28</p> <p>29</p> <p>30</p> <p>31</p> <p>32</p> <p>33</p> <p>34</p> <p>35</p> <p>36</p> <p>37</p> <p>38</p> <p>39</p> <p>40</p> <p>41</p> <p>42</p> <p>43</p> <p>44</p> <p>45</p> <p>46</p> <p>47</p> <p>48</p> <p>49</p> <p>50</p> <p>51</p> <p>52</p> <p>53</p> <p>54</p> <p>55</p> <p>56</p> <p>57</p> <p>58</p> <p>59</p> <p>60</p> <p>61</p> <p>62</p> <p>63</p> <p>64</p> <p>65</p> <p>66</p> <p>67</p> <p>68</p> <p>69</p> <p>70</p> <p>71</p> <p>72</p> <p>73</p> <p>74</p> <p>75</p> <p>76</p> <p>77</p> <p>78</p> <p>79</p> <p>80</p> <p>81</p> <p>82</p> <p>83</p> <p>84</p> <p>85</p> <p>86</p> <p>87</p> <p>88</p> <p>89</p> <p>90</p> <p>91</p> <p>92</p> <p>93</p> <p>94</p> <p>95</p> <p>96</p> <p>97</p> <p>98</p> <p>99</p> <p>100</p>	<p>1</p> <p>2</p> <p>3</p> <p>4</p> <p>5</p> <p>6</p> <p>7</p> <p>8</p> <p>9</p> <p>10</p> <p>11</p> <p>12</p> <p>13</p> <p>14</p> <p>15</p> <p>16</p> <p>17</p> <p>18</p> <p>19</p> <p>20</p> <p>21</p> <p>22</p> <p>23</p> <p>24</p> <p>25</p> <p>26</p> <p>27</p> <p>28</p> <p>29</p> <p>30</p> <p>31</p> <p>32</p> <p>33</p> <p>34</p> <p>35</p> <p>36</p> <p>37</p> <p>38</p> <p>39</p> <p>40</p> <p>41</p> <p>42</p> <p>43</p> <p>44</p> <p>45</p> <p>46</p> <p>47</p> <p>48</p> <p>49</p> <p>50</p> <p>51</p> <p>52</p> <p>53</p> <p>54</p> <p>55</p> <p>56</p> <p>57</p> <p>58</p> <p>59</p> <p>60</p> <p>61</p> <p>62</p> <p>63</p> <p>64</p> <p>65</p> <p>66</p> <p>67</p> <p>68</p> <p>69</p> <p>70</p> <p>71</p> <p>72</p> <p>73</p> <p>74</p> <p>75</p> <p>76</p> <p>77</p> <p>78</p> <p>79</p> <p>80</p> <p>81</p> <p>82</p> <p>83</p> <p>84</p> <p>85</p> <p>86</p> <p>87</p> <p>88</p> <p>89</p> <p>90</p> <p>91</p> <p>92</p> <p>93</p> <p>94</p> <p>95</p> <p>96</p> <p>97</p> <p>98</p> <p>99</p> <p>100</p>

## FEATURES

FEATURES	Location/Qualifiers
SOURCE	1. .859

```

I. 0.033
/organism="Zea mays"
/mol_type="genomic DNA"
/mol_strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBma068B11"
/clone_lib="ZM_0.7.1.5"
/note="vector: pBCK-;
methylation filtered;

```

## ORIGIN

	Query Match.	29.9%;	Score 569;	DB 29;	Length 859;
	Best Local Similarity	86.1%;	Pred. No. 3.1e-64;		
	Matches 691;	Conservative 0;	Mismatches 0;	Indels 112;	Gaps 17;
Qy	568	ATGAGGATGACGTGGACTCCATCTGCAAGGTTGGGTTCCGGGTCGAGATCGGCACGCTG	627		
Db	4	ATGAGGATGACGTGGACTCCATCTGCAAGGTTGGGTTCCGGGTCGAGATCGGCACGCTG	63		
Qy	628	TCGCCAGATCTCCCGAGAAACAGCTTCGCGCAGGCGTTCGATGCGGCCAACATCATCATC	687		
Db	64	TCGCCAGATCTCCCGAGAAACAGCTTCGCGCAGGCGTTCGATGCGGCCAACATCATCATC	123		
Qy	688	ACGTGCGGTTTCATCGAACCCGCTGTGCGGCATCAAGAGGTTCTTCCAGCTCGGTCAGAG	747		
Db	124	ACGTGCGGTTTCATCGAACCCGCTGTGCGGCATCAAGAGGTTCTTCCAGCTCGGTCAGAG	183		
Qy	748	GCCTCTCATGGCAGACATCMAGCTCGTGGACGAGTTCACCTACAGCGTGCATCGCCGG	807		
Db	184	GCCCTCTCTAGCGCAGAGCATCAAGCTCGTGGACGAGTTCACCTACAGCGTGCATCGCCGG	243		
Qy	808	AGGAAGCCCGAGATCGTCGAGGTCGGGGCCAGCGGCCAAACAGGAGA	853		
Db	244	AGGAAGCCCGAGATCGTCGAGGTCGGGGCCAGCGGCCAAACAGGAGAGGTACGTGCACAT	303		
Qy	854	-----	853		
Db	304	GACTGTTTCGATTCTTCAGTTTCATCGTCTTGGCGCGGATGACCTGATCTCTGATTGATTA	363		
Qy	854	-----	875		
Db	364	TATATCCATGTGACTTGTGAGGGCAATTAATTTGGCAGATGAGCACGACATCTCTGTC	423		
Qy	876	ACGGTTCATCGAGCTGGCGGAGGCGCGACGACGGCGCGGCTTCGGGGACGATAAGAG	935		
Db	424	ACGGTTCATCGAGCTGGCGGAGGCGCGGACGACGCGCGCGGCTTCGGGGACGATAAGAG	483		
Qy	936	CCTCCGGGACGTGTGTGTTCAACTTCGTGATCCCGCGGCGGACACGACGGCGGACGCT	995		

```

484 CCTCGGGAGCTGTGTCTAACTTCGTGATCCCGGGCGGACACGACGCGGCGCT 543
996 GTCTGTGTTACGACATGCGCATGTCCACCGGAGCTGTGCGGAGAGCTGCGCCGGA 1055
544 GTCTGTGTTACGACATGCGCATGTCCACCGGAGCTGTGCGGAGAGCTGCGCCGGA 603
1056 GCTGTGCGCGTTCGAGGCGGAGCGCGCGGAGAGGCGTCAAGCTCGTGTCTGCGG 1115
604 GCTGTGCGCGTTCGAGGCGGAGCGCGCGGAGAGGCGTCAAGCTCGTGTCTGCGG 663
1116 CGGCGTCAAGCGGAGCGGAGGCGGCGCGGCGGCGGAGGCGTCAAGCTCGTGTCTGCGG 1175
664 CGGCGTCAAGCGGAGCGGAGGCGGCGCGGCGGCGGAGGCGTCAAGCTCGTGTCTGCGG 723
1176 CACCTACGAGCGCTCGGAGGCGGAGGCGGCGGCGGCGGAGGCGTCAAGCTCGTGTCTGCGG 1235
724 CACCTACGAGCGCTCGGAGGCGGAGGCGGCGGCGGCGGAGGCGTCAAGCTCGTGTCTGCGG 783
1236 CCGTACCGCGCGCTCGGAGGCGGAGGCGGCGGCGGAGGCGTCAAGCTCGTGTCTGCGG 1258
784 CCGTACCGCGCGCTCGGAGGCGGAGGCGGCGGAGGCGTCAAGCTCGTGTCTGCGG 806

RESULT 7
AW519943 563 bp mRNA linear EST 06-MAR-2000
LOCUS 60053C10.Y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION AW519943
VERSION AW519943.1 GI:7162216
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 563)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL University
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660053 row: C column: 10.
Location/Qualifiers
1. -563
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XLOLR"
/clone_lib="660 - Mixed stages of anther and pollen"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."

FEATURES
source
Query Match 29.1%; Score 554; DB 10; Length 563;
Best Local Similarity 99.1%; Pred. No. 3e-62;
Matches 557; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORIGIN
854 AGATGAGCAGCAGCATCTGTACGTTTCAATCGAGCTGGCGGAGGCGCGGAGCGGCG 913
2 AGATGAGCAGCAGCATCTGTACGTTTCAATCGAGCTGGCGGAGGCGCGGAGCGGCG 61

```

```

914 GCCTCTTGGGGAGCATAGAGCTCCGGACCTGTGTGCTCAACTTGTGATCGCGGGC 973
62 GCCTCTTGGGGAGCATAGAGCTCCGGACCTGTGTGCTCAACTTGTGATCGCGGGC 121
974 GGCACACGCGGCGGAGCGCTGTGTGTTACGACATGGCCATGTCCCACTCCGAGC 1033
122 GGCACACGCGGCGGAGCGCTGTGTGTTACGACATGGCCATGTCCCACTCCGAGC 181
1034 TGGCCGAGAGAGCTTCGCGCGGAGCTGTGTGCGGTTTCGAGGCGGAGCGCGCGAGG 1093
182 TGGCCGAGAGAGCTTCGCGCGGAGCTGTGTGCGGTTTCGAGGCGGAGCGCGCGAGG 241
1094 GCGTCAAGCTGTGTGCTGTGCGGCGGCGCTGAGCGGCGGAGCAAGCGGTTTGGCGCGCGG 1153
242 GCGTCAAGCTGTGTGCTGTGCGGCGGCGCTGAGCGGCGGAGCAAGCGGTTTGGCGCGCGG 301
1154 TGGCGCAGTTTCGCGGCGGCTCCTCACCTACGACAGCTTCGCGCAAGCTGTGTGCTTCCCTCCACG 1213
302 TGGCGCAGTTTCGCGGCGGCTCCTCACCTACGACAGCTTCGCGCAAGCTGTGTGCTTCCCTCCACG 361
1214 CCGTCAAGCTGTGTGCTGTGCGGCGGCGCTGAGCGGCGGAGCAAGCGGTTTGGCGCGCGG 1273
362 CCGTCAAGCTGTGTGCTGTGCGGCGGCGCTGAGCGGCGGAGCAAGCGGTTTGGCGCGCGG 421
1274 TGGAGGACGAGCTGTGCTCCGCGGAGCGGAGCGGAGCTTCGCGGCGGAGCTGTGCTTCCCTCCACG 1333
422 TGGAGGACGAGCTGTGCTCCGCGGAGCGGAGCGGAGCTTCGCGGCGGAGCTGTGCTTCCCTCCACG 481
1334 TGGCGCAGTTTCGCGGCGGCTCCTCACCTACGACAGCTTCGCGCAAGCTGTGTGCTTCCCTCCACG 1393
482 TGGCGCAGTTTCGCGGCGGCTCCTCACCTACGACAGCTTCGCGCAAGCTGTGTGCTTCCCTCCACG 541
1394 CGGAGCGGTTGGATCAACGAGGA 1415
542 CGGAGCGGTTGGATCAACGAGGA 563

RESULT 8
CG047034 880 bp DNA linear GSS 19-AUG-2003
LOCUS PUFWN66TB ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta0749K12,
DEFINITION genomic survey sequence.
ACCESSION CG047034
VERSION CG047034.1 GI:33919214
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 880)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PUFWN66TD
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0206
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1. 880
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBta0749K12"

```

ORIGIN		/clone_lib="ZM_0.6.1.0_KB" /notes="Vector: pCR1-TOPO; Site 1: EcoRI; 0.6-1.0 kb high Cot selected genomic DNA library"	
Query Match Best Local Similarity 81.8%; Pred. No. 1.8e-60; Matches 720; Conservative 0; Mismatches 0; Indels 160; Gaps 2;			
QY	54	CCCACTAGCAGGCGCTCACAAGTACATCGCGCTCTCTGTTGTCTCTCATGGATCCT	113
Db	880	CCCACTAGCAGGCGCTCACAAGTACATCGCGCTCTCTGTTGTCTCTCATGGATCCT	821
QY	114	GGTCCAGAGTGGAGCTGAGGAAGCAGAAAGCCGAGATCATGCGCATCATCGCGC	173
Db	820	GGTCCAGAGTGGAGCTGAGGAAGCAGAAAGCCGAGATCATGCGCATCATCGCGC	761
QY	174	AACGGTGGACAGTGAAGAACTACACCGGATGCAGCATGGCTTGTGGGTACCTGTC	233
Db	760	AACGGTGGACAGTGAAGAACTACACCGGATGCAGCATGGCTTGTGGGTACCTGTC	701
QY	234	ACGGCAGACAGTGCCTGACATGCGCTTCACTTCTTACACCTACATCGCTGACCC	293
Db	700	ACGGCAGACAGTGCCTGACATGCGCTTCACTTCTTACACCTACATCGCTGACCC	641
QY	294	GGTGAATGTGAGCATGTCTCAAGATACTTACCAATTACCCCA	340
Db	640	GGTGAATGTGAGCATGTCTCAAGATACTTACCAATTACCCCAAGGTAATGACCT	581
QY	341	-----	340
Db	580	GAATCACTGATGTTTCACTCTTGGAAATCAGAGCTGAAAGTGAATGTCCTGA	521
QY	341	-----AGGAATCGTGTACAGATCTTACATGGAGCTGCTCTCGGTGACGGCATCT	391
Db	520	ACACCGTGTAGGAATCGTGTACAGATCTTACATGGAGCTGCTCTCGGTGACGGCATCT	461
QY	392	TCAACGCCGACGGAGCTGTGGAGGAAGCAGAGGAAGAGCGGAGTTTCAGTTCCCT	451
Db	460	TCAACGCCGACGGAGCTGTGGAGGAAGCAGAGGAAGAGCGGAGTTTCAGTTCCCT	401
QY	452	CCAAGAACCTGAGGATTTGAGCCCATTTGTTTCAGAGATCTTCTTGAAGCTGTGG	511
Db	400	CCAAGAACCTGAGGATTTGAGCCCATTTGTTTCAGAGATCTTCTTGAAGCTGTGG	341
QY	512	GTATATCAGCCAGCATCCAGGAGCAGCAAGTTTGGACATG	555
Db	340	GTATATCAGCCAGCATCCAGGAGCAGCAAGTTTGGACATG	281
QY	556	-----	555
Db	280	GGTCCCTTGCCATTGCCAATGAGCATTTCAA CCGTGAACACAGAGCTACCTTGGCGA	221
QY	556	--CAGGAATTTACATGAGGATGAGCTGGATCTCATCTGCAAGTTGGTTTCGGGTTCG	613
Db	220	TTCAAGAACCTTACATGAGGATGAGCTGGATCTCATCTGCAAGTTGGTTTCGGGTTCG	161
QY	614	AGATCGGACAGCTGTGCGCAGATCTCCCGAGAACAGCTTTCGCGCAGGCGTTTCGATGCGG	673
Db	160	AGATCGGACAGCTGTGCGCAGATCTCCCGAGAACAGCTTTCGCGCAGGCGTTTCGATGCGG	101
QY	674	CCAACATCATCATCGCTGCGGTTTCATCGACCCGCTGTGGCGCATCAAGAGTTTCCTCC	733
Db	100	CCAACATCATCATCGCTGCGGTTTCATCGACCCGCTGTGGCGCATCAAGAGTTTCCTCC	41
QY	734	ACGTGGGTTCAGAGCCCTCTTCAGCGCAGAGCATCAAGCT	773
Db	40	ACGTGGGTTCAGAGCCCTCTTCAGCGCAGAGCATCAAGCT	1

IP1\_20\_E12.bl\_A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,  
mRNA sequence.  
BI075441  
BI075441.1 GI:14514098  
EST.  
Sorghum bicolor (sorghum)  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 578)  
Klein, R., Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M. and  
Pratt, L.H.  
An EST database from Sorghum: developing preanthesis pannicles  
Unpublished (2001)  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for high quality sequence is  
20. Three-prime sequences, which are obtained with PolyMix or T7  
sequencing primer, are presented as the reverse complement.  
Seq primer: JEN REV  
High quality sequence stop: 576  
POLYA=No.

FEATURES  
Location/Qualifiers  
1..578  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultivar="BTx623"  
/db\_xref="taxon:4558"  
/clone\_lib="Immature pannicle 1 (IP1)"  
/note="Organ: Developing preanthesis pannicles; Vector:  
pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI;  
Site 2: EcoRI; The library was made from poly-A RNA in the  
cloning vector lambda ZAP II. Clones to be sequenced were  
prepared by mass excision."

ORIGIN  
Query Match 28.0%; Score 533.8; DB 12; Length 578;  
Best Local Similarity 96.5%; Pred. No. 1.2e-59;  
Matches 557; Conservative 0; Mismatches 17; Indels 3; Gaps 1;  
QY 488 GAGGATCTCCCTGGAAGCTGTGGGTATCTGAGCCAGGATCCAAAGCAGCAAGTTG 547  
Db 5 GAGGATCTCCCTGGAAGCTGTGGGTATCTGAGTCAAGCATCCAGGCGCAAGTTG 64  
QY 548 TGGACATSCAGGAACCTTACATGAGGATGAGCTGGACTCAATCTCAAGTTGGGTTG 607  
Db 65 TTGACATGCGAGGAACCTTACATGAGGATGAGCTGGACTCGAICTGCAAGTTGGGTTG 124  
QY 608 GGGTCGAGATCGGACGCTGTGCCAGATCTCCCGAGACAGCTTCGCCAGCGCTTCG 667  
Db 125 GGGTCGAGATCGGACGCTGTGCCGATCTTCCCGAGACAGCTTCGCCAGCGCTTCG 184  
QY 668 ATGCGCGCAACATCATCATCATCGCTGCGGTTTCATCGACCCGCTGTGGCGCATCAAGAGT 727  
Db 185 ATGCGCGTACATCATCATCGCTGCGGTTTCATCGACCCGCTGTGGCGCTCAAGAGT 244  
QY 728 TCTTCCACGCTGGGTCAAGGCCCTCTTAGCGCAGAGCATCAAGCTCGTGGACGAGTTCA 787  
Db 245 TCTTCCACGCTGGGTCAAGGCCCTCTTAGCGCAGAGCATCAAGCTCGTGGACGAGTTCA 304  
QY 788 CCTACAGGATCGCGCGGAGGAGCGCGAGATCTGTCAGGTCGCGGCGCGCAAC 847  
Db 305 CCTACAGGATCGCGCGGAGGAGCGCGAGATCTGTCAGGTCGCGGCGCGCAAC 364  
QY 848 AGGAGAGATCAAGCAGCATCTCTGTCAGGTTTCATCGAGCTGGGCGAGCGCGGACG 907



```

365 AGGAGAAGATGAAGACGACGACATCTCTGTCACGGTTCATCGAGCTGGCGGAGCGCGGACG 424
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

699 CCAATTACCCCAAGCTAAATGACCTGAACCTCACTGATGTTCACTGTTCTCGGAATCAGAGC 640
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

## RESULT 11

BI075273

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI075273 585 bp mRNA linear EST 20-JUN-2001  
 IP1\_18\_C12.bl\_A002 Immature pannicle 1 (IPI) Sorghum bicolor cDNA,  
 mRNA sequence.  
 BI075273 1 GI:14513930  
 EST.  
 Sorghum bicolor (sorghum)  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 585)  
 Klein, R.R., Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M. and  
 Pratt, J.H.  
 An EST database from Sorghum: developing preanthesis pannicles  
 Unpublished (2001)  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyTmix or T7 sequencing primer are presented as the reverse complement.

Seq primer: JEN REV

High quality sequence stop: 519

POLYA=No.

Location/Qualifiers

1. .585

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="BTx623"

/db\_xref="taxon:4558"

/clone\_lib="Immature pannicle 1 (IP1)"

/note="Organ: Developing preanthesis pannicles; Vector:

pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI;

Site 2: EcoRI; The library was made from poly-A RNA in the

cloning vector lambda ZAP II. Clones to be sequenced were

prepared by mass excision."

#### FEATURES

source

Query Match 26.7%; Score 509.4; DB 12; Length 585;

Best Local Similarity 93.6%; Pred. No. 1.7e-56;

Matches 544; Conservative 0; Mismatches 31; Indels 6; Gaps 1;

Qy 15 GGAAGCTCACCTCAGCGCGGCGAGCCATCGCCATTCCTCCACTAGCAGGGCTCAAA 74

Db 10 GGAAGCTCACCTCAGCGCGGCGA-----CACCATTGTTCCACTAGCAGGGCTCAAA 63

Qy 75 GTAATCGCGCTCTCTCTGTTGTTCTCTCTCATGATCTGTCTGACAGGTGAGCCTGAG 134

Db 64 GTACATCGCATCTCTCTCTGTTGTTCTCTCATGATGCGCTGTCTCATAGGTGAGCCTGAG 123

Qy 135 GAAGCAGAAAGCCGAGATCATGSCAGTATCGGCGCAACCGTGGAGCAGCTGAGGAA 194

Db 124 GAAGCAGAAAGCCGAGATCATGSCAGTATGCGCGCTATTGGCGCCACGTTGGAGCAGCTGAGGAA 183

Qy 195 CTACACCGGATGACGAGTGGCTTGTGCGGTACTGTGTCGCGCAGGACAGGACAGTACCGT 254

Db 184 CTACACCGGATGACGAGTGGCTTGTGCGGTACTGTGTCGCGCAGGACAGGACAGTACCGT 243

Qy 255 CGAATGCGCTTCTTCTTCTACACTATGCTGACCGGTGATGTCGAGCATGTCTCT 314

Db 244 CGAATGCGCTTCTTCTTCTTCTACACTATGCTGACCGGTGATGTCGAGCATGTCTCT 303

Qy 315 CAAGACTAACTTCAACAATTACCAAGGAACTCGTGTACAGATCCTTACATGGAAGTCT 374

Db 304 CAAGACTAACTTCAACAATTACCAAGGAACTCGTGTACAGATCCTTACATGGAAGTCT 363

Qy 375 CTTGGTGAAGGCATCTTCAAGCGCGAGCGGAGCTGTGGAGGAAGCAGAGAACGCGC 434

Db 364 CTTGGTGAAGGCATCTTCAAGCGCGAGCGGAGCTGTGGAGGAAGCAGAGAACGCGC 423

Qy 435 GAGTTTCAGTTCGCTCCAGAACTGAGGATTCAGGCCATTGTGTTTCAGAGATGA 494

Db 424 GAGTTTCAGTTCGCTCCAGAACTGAGGATTCAGTTCCAAATGTTTCAGAGATGA 483

Qy 495 CTCCTCGAAGCTGTGCGGTATATCTGAGCCAGGCATCCAAAGCAGGCAAAAGTTGTGACAT 554

Db 484 CTCCTCGAAGCTGTGCGGTATATCTGAGTCAGGCATCCAAAGCAGGCAAAAGTTGTGACAT 543

Qy 555 GCAGGAACCTTATCATGAGGATCAGCTGACATCTTGCA 595

Db 544 GCAGGAACCTTATCATGAGGATCAGCTGACATCTTGCA 584

#### RESULT 12

BE494080

LOCUS

DEFINITION WHE1277\_F02\_L032S Secale cereale anther cDNA library Secale cereale

CDNA clone WHE1277\_F02\_L03, mRNA sequence.

ACCESSION BE494080

VERSION BE494080.1

GI:9660673

KEYWORDS EST.

#### SOURCE

ORGANISM

Secale cereale (rye)  
Secale cereale  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Secale.

#### REFERENCE

AUTHORS

Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J.,  
Fenton, R.D., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y.,  
Lazo, G.R., Miller, R., Rausch, C.J., Ross, K., Seaton, C.L. and  
Tong, J.C.

#### TITLE

The structure and function of the expressed portion of the wheat

genomes - Anther cDNA library from rye

Unpublished (2000)

CONTACT: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: candershpw.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

#### FEATURES

source

1. .606

/organism="Secale cereale"

/mol\_type="mRNA"

/cultivar="Blanco"

/db\_xref="taxon:4550"

/clone="WHE1277\_F02\_L03"

/tissue\_type="Anther"

/dev\_stage="Adult plant before anthesis"

/lab\_host="E. coli SOLR"

/clone\_lib="Secale cereale anther cDNA library"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;

Site 1: EcoRI; Site 2: XhoI; plants were grown in the

greenhouse. Anthers were harvested and pooled from early

meiosis to late meiosis. The tissue, total RNA, and

poly(A) RNA were prepared (Butler, Ross and Gustafson) at

University of Missouri, Columbia. A cDNA library was

made, and the cDNA clones were in vivo excised to give

pBluescript phagemids in the TJ Close lab (Choi, Close,

Fenton) at the University of California, Riverside.

Plasmid DNA preparations and DNA sequencing were performed

in the OD Anderson lab (all other authors)."

#### ORIGIN

Query Match 25.3%; Score 482; DB 10; Length 606;

Best Local Similarity 87.5%; Pred. No. 6e-53;

Matches 527; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 988 ACGACGCTGTCTGTTTACGCATGGCCATGTCCACCCGAGCTGCGCGAGAGCTG 1047

Db 5 ACGACGCTGTCTGTTTACGCATGGCCATGTCCACCCGAGCTGCGCGAGAGCTG 64

Qy 1048 CGCGCGAGCTGTGCGGCTTCGAGCGGAGCGCGCGCGAGGAGGCGTCCAGCTCGT 1107

Db 65 CGCGCGAGCTGTGCGGCTTCGAGCGGAGCGCGCGCGAGGAGGCGTCCAGCTCGT 124

Qy 1108 CTCTGCGCGCGCTGTGACCGCGACGCAAGAGCGTTCGCGCGCGCGCTGGCGAGTTCGCG 1167

Db 125 CCTGTGACGAGCGGAGGGGTTCGACGAGGCGCTTCGCGCGCGCGCTGGCGAGTTCGCG 184

Qy 1168 GGCCTCTCTACCTACGACAGCTCGGCAAGTGTGTACTCTCCAGCGCTGGTCCACGAG 1227

Db 185 GGCCTCTCTACCTACGACAGCTCGGCAAGTGTGTACTCTCCAGCGCTGGTCCACGAG 244

Qy 1228 AGCTTCGCTGTATCCCGCGCTTCCTCAGACCCAGGGATCTCTGAGAGCAGCTG 1287

Db 245 AGCTTCGCTGTATCCCGCGCTTCCTCAGACCCAGGGATCTCTGAGAGCAGCTG 304

Qy 1288 CTGCGGAGCGGACGAGAGGTGAGGCGCGCGGAGTGGTGAGCTAGTCCCTTACTTCGATG 1347



```
source
1. .517
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XfOLR"
/clone_lib="660 - Mixed stages of anther and pollen"
/notes="Organ: anthers; Vector: Lambda Zap; Site:1: EcoRI; Site:2: XhoI; Anther and pollen cDNA library. Directionally sequenced with 5' end at the EcoRI site. Created by Amie Franklin."

ORIGIN
Query Match 24.3%; Score 463.6; DB 10; Length 517;
Best Local Similarity 94.6%; Pred. No. 1.5e-50;
Matches 510; Conservative 0; Mismatches 5; Indels 24; Gaps 2;

Qy 1296 CCGGACGAAGTGTAGGCGCGGGATGGTGCAGTACGTGCGCTTCTCGATGGGCGGAT 1355
Db 515 CCGGACGAAGTGTAGGCGCGGGATGGTGCAGTACGTGCGCTTCTCGATGGGCGGAT 456
Qy 1356 GGAGTACAACCTGGGCGCGCGGAGCTTCCGCGGAGCGGTGGATCAACGAGGA 1415
Db 455 GGAGTACAACCTGGGCGCGCGGAGCTTCCGCGGAGCGGTGGATCAACGAGGA 396
Qy 1416 TGGCGCGTTCGCAACGGTTCGCGTTCACGTTACCGCGTTCAGCGCGGCGGAGGAT 1475
Db 395 TGGCGCGTTCGCAACGGTTCGCGTTCACGTTACCGCGTTCAGCGCGGCGGAGGAT 336
Qy 1476 CTGCTTGGCAAGGACTCGGCGTACCTGCAGATGAAGATGCGGTGGCGTTCCTTCGCG 1535
Db 335 CTGCTTGGCAAGGACTCGGCGTACCTGCAGATGAAGATGCGGTGGCGTTCCTTCGCG 276
Qy 1536 CTTTACAGCTTCGGCTGTCTGGAGGGGACACCGGTGCAGTACCGCATGATGACCATCT 1595
Db 275 CTTTACAGCTTCGGCTGTCTGGAGGGGACACCGGTGCAGTACCGCATGATGACCATCT 216
Qy 1596 CTCCATGGCGCAGCGCTCAAGGTCGCGTCTTAGGCGCTCTGATGTCATGGCGATTT 1655
Db 215 CTCCATGGCGCAGCGCTCAAGGTCGCGTCTTAGGCGCTCTGATGTCATGGCGATTT 156
Qy 1656 GGATATGATATCGTCCCGCTTAATCCACGAAATACGCTCGGTGTACAAATTTGCAT 1715
Db 155 G-----GGATATATCCCGCTTAATCC-----TTAAAAATTTGCAT 120
Qy 1716 GCATGATGTAAGGAAGGATGGTTTCAATGTGTGGCTTGGCTTAAGCCTTAAAAACT 1775
Db 119 GCATGATGTAAGGAAGGATGGTTTCAATGTGTGGCTTGGCTTAAGCCTTAAAAACT 60
Qy 1776 CCGTCGGCTTGGGAACACACATCACTAGTGTGTTTGTACTCTACTCTCAGTGGAA 1834
Db 59 CCGTCGGCTTGGGAACACACATCACTAGTGTGTTTGTACTCTACTCTCAGTGGAA 1

RESULT 15
BQ841568 591 bp mRNA linear EST 12-AUG-2002
LOCUS WHE4212_E08_J16ZS Aegilops speltoides pre-meiotic anther cDNA
DEFINITION library Aegilops speltoides cDNA clone WHE4212_E08_J16, mRNA
sequence.
ACCESSION BQ841568
VERSION BQ841568.1 GI:22210977
KEYWORDS EST
SOURCE Aegilops speltoides
ORGANISM Aegilops speltoides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Aegilops.
REFERENCE 1 (bases 1 to 591)
AUTHORS Akhunov,E., Anderson,O.D., Chao,S., Crossman,C., Devorak,J.,
Lazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
```

```
TITLE The structure and function of the expressed portion of the wheat
JOURNAL genomes - Pre-meiotic anther cDNA library from Aegilops speltoides
COMMENT Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: canders@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES
Location/Qualifiers
1. .591
/organism="Aegilops speltoides"
/mol_type="mRNA"
/cultivar="F2 from 2-12-4-8-1-1-1-(1) x
P136909-12-811-(1)"
/db_xref="taxon:4573"
/clone="WHE4212_E08_J16"
/tissue_type="Anther"
/dev_stage="pre-meiotic anthers"
/lab_host="E. coli SOLR"
/clone_lib="Aegilops speltoides pre-meiotic anther cDNA
library"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in a growth
chamber at the University of California, Davis (Akhunov in
Devorak's lab). Pre-meiotic anthers were harvested, total
RNA and poly(A) RNA were prepared, a cDNA library was
made, and the cDNA clones were in vivo excised to give
pBluescript phagemids at the University of California,
Davis (Akhunov in Devorak's lab). Plasmid DNA preparations
and DNA sequencing were performed in the OD Anderson lab
(all other authors)."
```

Qy	5	TCGGCAGAGGAGCTCCTCAGCGCGGAGCCATCGCCATTTCTCCCATAGCAG	64
Db	1	TCGGCAGAGGAGCTCCTTGGCATCGCTGACGCGGCTTCTCCCGTGGCAG	60
Qy	65	GGCCTCAAGTATCATCGCGCTCTTCTGTTGTCTCTCATGATCTGTGTCAGAGGT	124
Db	61	GGCTCCCAAGTTCATGCGCATCTTCTGTTGTCTCTGTTGATCTGTGTCACCTGT	120
Qy	125	GGAGCCTTGAGGAAGCAGAAAGGCGCGAGATCATGCGCAGTCATCGCGCAACGGTGGAGC	184
Db	121	GGAGCCTTGAGGAAGCAGAAAGGCGCGAGGTCTATGGCGGTCATCGCGCCACGCTGGAGC	180
Qy	185	AGCTGAGGAATACACCGGATGACGACTGGCTTGTGGGTACTGTACGGGACACGGA	244
Db	181	AGCTGAGGAATACTACCGGATGACGACTGGCTCGTGGAGTACTGTCCAAGCACCGGA	240
Qy	245	CAGTGACCGTCGATCGCTTCACTTCTTACCTACCTACATCGCTGACCGGTGAATGTGCG	304
Db	241	CGGTACCGTCGATCGCTTCACTTCTTACCTTACCTACATCGCTGACCGGTGAATGTGCG	300
Qy	305	AGCATGTCTCAAGACTAATTTACCAATTAACCCCAAGGGGAATCGTGTACAGATCTTACA	364
Db	301	AGCATGTCTCAAGACTAATTTCAACAATTAACCCCAAGGGGAATCGTGTACAGATCTTACA	360
Qy	365	TGGAGCTGCTCTCGGTGACGCGATCTTCAACGCGCGGCGGAGCTGTGGAGGAGCAGA	424
Db	361	TGGAGCTGCTCTCGGTGACGCGATTTCAACGCGCGGCGGAGCTGTGGAGGAGCAGA	420
Qy	425	GGAGACGGCGAGTTTCGAGTTTCGCTCCCAAGAACCTGAGGATTTTCAGGCGCATTTGTGT	484
Db	421	GGAGACGGCGAGTTTCGAGTTTCGCTCCCAAGAACCTGAGGATTTTCAGGCGCATTTGTGT	480

Query Match 24.2%; Score 461.4; DB 13; Length 591;  
Best Local Similarity 86.3%; Pred. No. 2.8e-50;  
Matches 510; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

us-10-021-657-1.rst

Mon Mar 8 10:13:27 2004

Qy	485	TCAGAGAGTACTCCCTGAAGCTGTGGGTATACCTGAGCCAGGCATCCAAAGGCGGCAAG	544
Db	481	TCCGGGAGTACTCCCTGAAGCTGTCCAGCATCCTGAGCCAGGCTTGCAAGGCCGCGCAAG	540
Qy	545	TTGTGGACATGCAGGAACCTTACATGAGGATGACGCTGGACTCCATCTGCA	595
Db	541	TTGTGGACATGCAGGAGCTGTACATGAGGATGACGCTGGACTGATCTGCA	591

Search completed: March 6, 2004, 02:18:15  
Job time : 5032 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2004, 00:53:15 ; Search time 678 Seconds  
(without alignments)  
10281.808 Million cell updates/sec

Title: US-10-021-657-1

Perfect score: 1906

Sequence: 1 gaattggcagcgagggaagc.....aaaaaaaaaaactcgag 1906

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2421054 seqs, 1828716029 residues

Total number of hits satisfying chosen parameters: 4842108

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1906	100.0	1906	13	US-10-021-657-1
2	1906	100.0	1906	14	US-10-412-000-1
3	1378.4	72.3	3897	13	US-10-021-657-7
4	1378.4	72.3	3897	14	US-10-412-000-7
5	433.2	22.7	494	13	US-10-021-657-3
6	433.2	22.7	494	14	US-10-412-000-3
7	392.4	20.6	1807	12	US-10-425-114-23429
8	298.2	15.6	1847	12	US-10-425-114-28465
9	289	14.1	1617	15	US-10-260-238-1236
10	255	13.4	1694	12	US-10-425-114-28590
11	228	12.0	1638	15	US-10-260-238-907
12	226.4	11.9	1527	15	US-10-260-238-422
13	217.2	11.4	1680	12	US-10-425-114-1524
14	206	10.8	1806	12	US-10-425-114-28220
15	198	10.4	1761	12	US-10-425-114-26392

16 194.2 10.2 825 12 US-10-425-114-35451 Sequence 35451, A  
17 183.8 9.6 1498 12 US-10-425-114-17454 Sequence 17454, A  
18 178 9.3 1267 12 US-10-425-114-32933 Sequence 32933, A  
19 166.2 8.7 854 15 US-10-260-238-4619 Sequence 4619, Ap  
20 165.4 8.7 1473 9 US-09-938-842A-1981 Sequence 1981, Ap  
21 165.4 8.7 1473 11 US-09-938-842A-1981 Sequence 1981, Ap  
22 163.6 8.6 654 15 US-10-260-238-4618 Sequence 4618, Ap  
23 162.2 8.5 658 15 US-10-260-238-4017 Sequence 4017, Ap  
24 161.6 8.5 444 15 US-10-260-238-5268 Sequence 5268, Ap  
25 160.4 8.4 792 15 US-10-260-238-4608 Sequence 4608, Ap  
26 159 8.3 2431 12 US-10-424-599-120032 Sequence 120032, A  
27 157.4 8.3 2239 12 US-10-425-114-10296 Sequence 10296, A  
28 149.4 7.8 604 15 US-10-260-238-3255 Sequence 3255, Ap  
29 148.2 7.8 1682 9 US-09-938-842A-881 Sequence 881, Ap  
30 148.2 7.8 1662 11 US-10-424-599-20189 Sequence 20189, A  
31 136.8 7.2 1462 12 US-10-424-599-2678 Sequence 2678, Ap  
32 135.6 7.1 1614 9 US-09-938-842A-2678 Sequence 2678, Ap  
33 135.6 7.1 1614 11 US-09-938-842A-2678 Sequence 120026, A  
34 134.8 7.1 2763 12 US-10-424-599-120026 Sequence 21, Appl  
35 131.8 6.9 1888 15 US-10-429-949-21 Sequence 4018, Ap  
36 129.8 6.8 520 15 US-10-260-238-4018 Sequence 103097, A  
37 128.2 6.7 756 12 US-10-424-599-103097 Sequence 572, Ap  
38 127.6 6.7 620 12 US-10-425-114-4749 Sequence 1868, Ap  
39 120.4 6.3 866 9 US-09-770-445-572 Sequence 1868, Ap  
40 119 6.2 1542 9 US-09-938-842A-1868 Sequence 1868, Ap  
41 119 6.2 1542 11 US-09-938-842A-1868 Sequence 880, Ap  
42 118.2 6.2 1729 12 US-10-425-114-880 Sequence 17485, A  
43 116.8 6.1 1872 12 US-10-424-599-17485 Sequence 32508, A  
44 114 6.0 1199 12 US-10-425-114-32508 Sequence 120029, A  
45 107.6 5.6 1349 12 US-10-424-599-120029

#### ALIGNMENTS

RESULT 1  
US-10-021-657-1  
; Sequence 1, Application US/10021657  
; Publication No. US00020083483A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; TITLE OF INVENTION: METHOD OF USING SAME  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021.657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1906  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1..1638, 1642..1767)  
US-10-021-657-1

Query Match 100.0%; Score 1906; DB 13; Length 1906;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTGGCAGGAGGAAGCTCACTCAGCGGCGAGCGCATGCCATCTTCCACTA 60  
DB 1 GAATTGGCAGGAGGAAGCTCACTCAGCGGCGAGCGCATGCCATCTTCCACTA 60  
QY 61 GCAGGGCTCACAAGTACATCGCGTCTTCTGTTGCTCTCATGATCTGTGTCAG 120  
DB 61 GCAGGGCTCACAAGTACATCGCGTCTTCTGTTGCTCTCATGATCTGTGTCAG 120  
QY 121 AGGTGGAGCTGAGGAAGCAGAAAGCCCGAGATCATGCCAGTCATCGGCGCAACGGTG 180

121 AGGTGGAGCCTTGAGGAGCAGAGGAGCCCGAGATCATGGCCAGTCACTCGGCGCAACGGT 180  
181 GAGCAGCTGAGGAACTACCAACGGATGACAGCTGGCTTGTGCGGTACTGTCTACGGCAC 240  
181 GAGCAGCTGAGGAACTACCAACGGATGACAGCTGGCTTGTGCGGTACTGTCTACGGCAC 240  
241 AGGACAGTGCCTGAGACATGCGCTTCACTTCTACCTACCTACCTACCTACCTACCTACCT 300  
241 AGGACAGTGCCTGAGACATGCGCTTCACTTCTACCTACCTACCTACCTACCTACCTACCT 300  
301 GTCGAGCAGTGTCTCTCAAGACTAACTTCAACCAATACCCCAAGGAAATCGTGTACAGATCC 360  
301 GTCGAGCAGTGTCTCTCAAGACTAACTTCAACCAATACCCCAAGGAAATCGTGTACAGATCC 360  
361 TACATGAGCAGTGTCTCTGAGTGCAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 420  
361 TACATGAGCAGTGTCTCTGAGTGCAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 420  
421 CAGAGGAGCAGCGGAGTTCGAGTTCGCTTCAAGAACTCAGGAGATTCAGCGCCATT 480  
421 CAGAGGAGCAGCGGAGTTCGAGTTCGCTTCAAGAACTCAGGAGATTCAGCGCCATT 480  
481 GTGTTTCAAGAGTACTTCCCTGAGCTGTGCGGTATCTAGCGAGCATCCCAAGGAGGC 540  
481 GTGTTTCAAGAGTACTTCCCTGAGCTGTGCGGTATCTAGCGAGCATCCCAAGGAGGC 540  
541 AAGTTTGTGACATGACAGAACTTTACATGAGATGACCTGAGCTTCAAGGTT 600  
541 AAGTTTGTGACATGACAGAACTTTACATGAGATGACCTGAGCTTCAAGGTT 600  
601 GGGTTCGAGTGCAGATCGGAGTTCGCTTCAAGAACTCAGGAGATTCAGCGCCATT 660  
601 GGGTTCGAGTGCAGATCGGAGTTCGCTTCAAGAACTCAGGAGATTCAGCGCCATT 660  
661 GCGTTTCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 720  
661 GCGTTTCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 720  
721 AAGAGTTTTCACGTCGAGTGCAGGAGTTCAGGAGTTCAGGAGTTCAGGAGTTCAGGAGTTC 780  
721 AAGAGTTTTCACGTCGAGTGCAGGAGTTCAGGAGTTCAGGAGTTCAGGAGTTCAGGAGTTC 780  
781 GAGTTTCACTTACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
781 GAGTTTCACTTACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
841 GGCAACAGGAGAGATGAGAGCAGATCATCTGTACGGTTTCACTGAGCTGGGCGAGGCC 900  
841 GGCAACAGGAGAGATGAGAGCAGATCATCTGTACGGTTTCACTGAGCTGGGCGAGGCC 900  
901 GCGGACAGCGGCGGCTTGGGAGAGATGAGAGCTTCGGGAGCTGGGAGCTGGGAGCTGGGAG 960  
901 GCGGACAGCGGCGGCTTGGGAGAGATGAGAGCTTCGGGAGCTGGGAGCTGGGAGCTGGGAG 960  
961 GTGATGCGCGGCGGAGACAGCGGAGCTGCTGTGTTTACGCGCATGGCCATGGCCATG 1020  
961 GTGATGCGCGGCGGAGACAGCGGAGCTGCTGTGTTTACGCGCATGGCCATGGCCATG 1020  
1021 TCCACCGGAGCTGGGCGGAGAGCTGGGCGGAGCTGGGCGGAGCTGGGCGGAGCTGGGCGG 1080  
1021 TCCACCGGAGCTGGGCGGAGAGCTGGGCGGAGCTGGGCGGAGCTGGGCGGAGCTGGGCGG 1080  
1081 GCGGCGGAGGAGGCTTCAAGCTTCTGCGGCGGCTGAGCGGAGCAAGGCG 1140  
1081 GCGGCGGAGGAGGCTTCAAGCTTCTGCGGCGGCTGAGCGGAGCAAGGCG 1140  
1141 TTGCGGCGGCGGCTGGGCGGAGTTCGCGGCGCTTCTTACCTACGAGCTTCGCGGAGCTG 1200  
1141 TTGCGGCGGCGGCTGGGCGGAGTTCGCGGCGCTTCTTACCTACGAGCTTCGCGGAGCTG 1200  
1201 GTCTACCTCCAGCGCTGCTGATCCAGAGAGCTTCGCGCTGATCCCGCGCTTCTCTCAGGAC 1260

1201 GTCTACCTCCAGCGCTGCTGATCCGAGACGCTCGGCTGTACCCGCGCTTCTCTCAGGAC 1260  
1261 CCCAAGGGGATCTCTGGAGGACGACGCTGCTCCGGAACGAGGAGTGGAGGCGGCGGG 1320  
1261 CCCAAGGGGATCTCTGGAGGACGACGCTGCTCCGGAACGAGGAGTGGAGGCGGCGGG 1320  
1321 ATGTGAGCTACGTCGCTTACTGATGGGCGGAGTGCAGTACACTGGGCGGCGGAGCGG 1380  
1321 ATGTGAGCTACGTCGCTTACTGATGGGCGGAGTGCAGTACACTGGGCGGCGGAGCGG 1380  
1381 GCGAGCTTCCGCGCGGAGCGGTGATCAACGAGGATGCGGCTTCCGCAACGCGCTCGCG 1440  
1381 GCGAGCTTCCGCGCGGAGCGGTGATCAACGAGGATGCGGCTTCCGCAACGCGCTCGCG 1440  
1441 TTCAAGTTTACGCGCTTCCAGCGCGGCGGAGGATCTGCTGGGCAAGGATCTCGGCGTAC 1500  
1441 TTCAAGTTTACGCGCTTCCAGCGCGGCGGAGGATCTGCTGGGCAAGGATCTCGGCGTAC 1500  
1501 CTGCAGATGAGATGGCGCTGGCCATCTCTTCCGCTTCTACAGCTTCCGCGCTCTCGAG 1560  
1501 CTGCAGATGAGATGGCGCTGGCCATCTCTTCCGCTTCTACAGCTTCCGCGCTCTCGAG 1560  
1561 GGGCACCGGCTGCAGTACCGCATGATGATGATGATGATGATGATGATGATGATGATGATG 1620  
1561 GGGCACCGGCTGCAGTACCGCATGATGATGATGATGATGATGATGATGATGATGATGATG 1620  
1621 GCGCTTCTAGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680  
1621 GCGCTTCTAGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680  
1681 CCACGACAAATACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740  
1681 CCACGACAAATACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740  
1741 GTTTCATTTAGGCTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1800  
1741 GTTTCATTTAGGCTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1800  
1801 TCACTAGTGTGTTGATCTTACT 1860  
1801 TCACTAGTGTGTTGATCTTACT 1860  
1861 ATATATATATCT 1906  
1861 ATATATATATCT 1906

RESULT 2

US-10-412-000-1  
; Sequence 1, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: METHOD OF USING SAME  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1906  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1..1638, 1642..1767)  
US-10-412-000-1









2528 CTAACGACGCTCGGACGCTGGTCTACCTCCACGCTGCTGCTACCGAGCCTCCGCT 2587  
1239 GTACCCCGCGCTCCCT-----1254  
2588 GTACCCCGCGCTCCCTCAGGTGAGCGCGCCGACGACGACCTCCGGTCCAGACACAGC 2647  
1255 -----CAGGACCCCAAGG 1267  
2648 ATGCAGTGAGTGACCTGATGCAATGACATGCACTTGCGCGCGCGCAGGACCCCAAGG 2707  
1268 GGATCTCGAGGACGACGCTGCTCCCGACGCGGACGAGTGAGGCGCGCGGATGGTGA 1327  
2708 GGATCTCGAGGACGACGCTGCTCCCGACGCGGACGAGTGAGGCGCGCGGATGGTGA 2767  
1328 CGTACCTGCGCTACTCGATGGGCGGATGAGTACAACTGGGCGCCCGACGCGCGGAGCT 1387  
2768 CGTACCTGCGCTACTCGATGGGCGGATGAGTACAACTGGGCGCCCGACGCGCGGAGCT 2827  
1388 TCCGCGCGGACGCGTGGATCAACGAGGATGGCGGCTTCCGCAACGCGTCCGCTTCAAGT 1447  
2828 TCCGCGCGGACGCGTGGATCAACGAGGATGGCGGCTTCCGCAACGCGTCCGCTTCAAGT 2887  
1448 TCACGGGCTTCAGGCGCGGCGGAGGATCTGCTGGCAAGGACTCGGCTACCTGCAGA 1507  
2888 TCACGGGCTTCAGGCGCGGCGGAGGATCTGCTGGCAAGGACTCGGCTACCTGCAGA 2947  
1508 TGAAGATGGCGCTGGCGCATCTCTTCCGCTTCTACAGCTTCCGCTGCTGAGGGGCAAC 1567  
2948 TGAAGATGGCGCTGGCGCATCTCTTCCGCTTCTACAGCTTCCGCTGCTGAGGGGCAAC 3007  
1568 CGGTGCAAGTACCGCATGATGACCATCTCTCCATGGCGCGGCTCAAGGTCCGCTCT 1627  
3008 CGGTGCAAGTACCGCATGATGACCATCTCTCCATGGCGCGGCTCAAGGTCCGCTCT 3067  
1628 CTAGGCGCTGATGATGAGGATTTGGATATGGATATGCTCCGCTTAACTCCAGC 1687  
3068 CTAGGCGCTGATGATGAGGATTTGGATATGGATATGCTCCGCTTAACTCCAGC 3117  
1688 AATAAGCTCGTCTTACAAATTTGCATGATGATGATGATGATGATGATGATGATGAT 1747  
3118 -----TTAAATTTGCATGATGATGATGATGATGATGATGATGATGATGAT 3163  
1748 TGGTGGCTTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAA 1807  
3164 TGGTGGCTTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAA 3223  
1808 TGGTGGCTTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAA 1867  
3224 TGGTGGCTTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAA 3283  
1868 TTATCCTCTTCTT 1891  
3284 TTATCCTCTTCTT 3297

RESULT 5  
US-10-021-657-3  
; Sequence 3, Application US/10021657  
; Publication No. US20020083483A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021,657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 3

LENGTH: 494  
TYPE: DNA  
ORGANISM: Sorghum sp.  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (1)..(494)  
OTHER INFORMATION: "n" bases may be a, t, c, g, other or unknown  
US-10-021-657-3

Query Match 22.7%; Score 433.2; DB 13; Length 494;  
Best Local Similarity 93.7%; Pred. No. 2.4e-106;  
Matches 447; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 259 ATGCGGTTTCACTTCTTACACCTTACCTGCTGACCCCGGTGAATGTCGAGCATGCTCTCAAG 318  
DB 13 ATGCGGTTTCACTTCTTACACCTTACCTGCTGACCCCGGTGAATGTCGAGCATGCTCTCAAG 72  
QY 319 ACTAACTTCCACCAATTAACCCAAAGGAATTCGCTGATCAGATCCTTACATGAGCATGCTCTC 378  
DB 73 ACTAACTTCCACCAATTAACCCAAAGGAATTCGCTGATCAGATCCTTACATGAGCATGCTCTC 132  
QY 379 GTGACGGCATTTTCAACCGCGCGGAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 438  
DB 133 GTGACGGCATTTTCAACCGCGCGGAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGT 192  
QY 439 TTCGAGTTTCCCTTCCAAAGAACTTCCGCGGATTTTCCAGCCCATTTGTTTCCAGAGTACTCC 498  
DB 193 TTCGAGTTTCCCTTCCAAAGAACTTCCGCGGATTTTCCAGCCCATTTGTTTCCAGAGTACTCC 252  
QY 499 CTGAAGCTTCCGCTTCCAAAGAACTTCCGCGGATTTTCCAGCCCATTTGTTTCCAGAGTACTCC 558  
DB 253 CTGAAGCTTCCGCTTCCAAAGAACTTCCGCGGATTTTCCAGCCCATTTGTTTCCAGAGTACTCC 312  
QY 559 GAATTTTACATGAGGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 618  
DB 313 GAATTTTACATGAGGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 372  
QY 619 GGCACGCTTCCGCTTCCAAAGAACTTCCGCGGATTTTCCAGCCCATTTGTTTCCAGAGTACTCC 678  
DB 373 GGCACGCTTCCGCTTCCAAAGAACTTCCGCGGATTTTCCAGCCCATTTGTTTCCAGAGTACTCC 432  
QY 679 ATCATCATCAGCTGCGGTTTCACTGACCCCGCTGTGGGCGATCAAGAGGTTCTTCCAC 735  
DB 433 ATCATCTGCTACNCTGCGGTTTCACTGACCCCGCTGTGGGCGATCAAGAGTACTTCTTCCAC 489

RESULT 6  
US-10-412-000-3  
; Sequence 3, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 3  
LENGTH: 494  
TYPE: DNA  
ORGANISM: Sorghum sp.  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (1)..(494)  
OTHER INFORMATION: "n" bases may be a, t, c, g, other or unknown  
US-10-412-000-3

Query Match 22.7%; Score 433.2; DB 14; Length 494;  
Best Local Similarity 93.7%; Pred. No. 2.4e-106;  
Matches 447; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

259 ATGCGGTTCTTCTTACACCTACATCGCTGACCCGGTGAATGTCGAGCATGTCCTCAAG 318  
13 ATGCGGTTCTTCTTACACCTACATCGCTGACCCGGTGAATGTCGAGCATGTCCTCAAG 72  
319 ACTAACTTCCACCAATTACCCCAAGGAATCGTGTACAGATCCTACATGAGCGTGTCTCTC 378  
73 ACTAACTTCCACCAATTACCCCAAGGAATCGTGTACAGATCCTACATGAGCGTGTCTCTC 132  
379 GGTGACGGCATCTTCAACGCGGACGCGGAGCTGTGGAAGGAAGAGAGAGAGAGAGAGT 438  
133 GGTGACGGCATCTTCAACGCGTGTACGCGGAGCTGTGGAAGGAAGAGAGAGAGAGAGT 192  
439 TTCAGATTGCGCTTCAAGAACTTACGAGGATTCAGCGCCATTTGTTTCAGAGATCTCC 498  
193 TTCAGATTGCGCTTCAAGAACTTACGAGGATTTCAAGTCCCAATGTTTTCAGAGATCTCC 252  
499 CTGAAGCTGTGCGGTATCTAGCGAGCATCCCAAGGAGGCAAGTTGTGGAATCGAG 558  
253 CTGAAGCTGTGCGGTATCTAGCGAGCATCCCAAGGAGGCAAGTTGTGGAATCGAG 312  
559 GAACTTTACATGAGATGACCTGAGTCTCAATCTGCAAGTTGGGTTGGGTTGAGATC 618  
313 GAACTTTACATGAGATGACCTGAGTCTCAATCTGCAAGTTGGGTTGGGTTGAGATC 372  
619 GGCAGCTGTGCGGATCTCCCGAGAACAGCTTCGCGAGGCTTCGATGCGGCCAAC 678  
373 GGCAGCTGTGCGGATCTCCCGAGAACAGCTTCGCGAGGCTTCGATGCGGCCAAC 432  
679 ATCATCATCAGCTGCGGTTTCATCGAACCCGCTGTGCGGATCAAGAGGTTCTTCCAC 735  
433 ATCATCATCAGCTGCGGTTTCATCGAACCCGCTGTGCGGATCAAGAGGTTCTTCCAC 489

RESULT 7  
US-10-425-114-23429  
; Sequence 23429, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yinhua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 23429  
; LENGTH: 1807  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3596-027-C6\_FLI  
US-10-425-114-23429

Query Match 20.6%; Score 392.4; DB 12; Length 1807;  
Best Local Similarity 58.2%; Pred. No. 3.8e-95;  
Matches 855; Conservative 0; Mismatches 531; Indels 84; Gaps 6;

159 GCAGTTCATCGGGGCAACGGTGGAGAGCTGAGGAATACCAAGGATGACAGACTGGCT 218  
179 GCGGTTGGTGGGACGGTGTTCACCAAGTGTACCAAGTGTACCAAGTGTTCGCGGCTGACAGACTACCA 238  
219 TGTGCGGTACTCTTACCGGACAGGACAGTACCGTTCGATCGATCGCG---TTCACTTCTCTA 275

Db 239 CACGAGCTGTTCGCGCAGCGCAAGACCTTCCAGCTGTGTCGCCGCCGCCGCCGCCGCA 298  
QY 276 CACCTTACATCGCTGACCCCGTGAATGTCGAGCATGTCCTCAAGACTAACTTCAACCAATTA 335  
Db 299 GATATACACGTGCGACCCCGCGTGGTGCAGACATCTCTCAGGACCACTTTCGCCAATTA 358  
QY 336 CCCAAGGGAATCGTGTACAGATCTCTACATGACGTCCTCGGTGAGCGGCAATTTCAA 395  
Db 359 CGCAAGGGCAGGTTCAACTACGAGAACATGACTGACCTGCTCGCGGAGCGGATCTTCGC 418  
QY 396 CGCGCAGCGGAGCTGTGAGGAAGCAGAGGAAGAGCGGAGTTTCGAGTTCGCTCCCA 455  
Db 419 CGTGGACGCGCAACAGTGGAGGAGCAGCGGGAAGATCGCCAGCTACGACTTCTTCCACGAG 478  
QY 456 GAACCTGAGGATTTTCAGCGCCATTTGTTTCAGAGAGTATCTCCCTCAAGCTGTGCGGTAT 515  
Db 479 GCGCTGCGCGATTTTCAGCGGCGCGTCTTCAAGAGAAAGCGCGGAGGCTCGCGGCAT 538  
QY 516 ACTGAGCCAGGCATCCAAAGGACGCGCAAGTTGTGGAACATGCGAGAACTTTTACATGAGGAT 575  
Db 539 CGTCTCGGCAACGCGGCGTTCGCGGAGCCCATGGAGTTCCAGGGCTCGCGCTCAGGGC 598  
QY 576 GACGCTGGAATCTTCGCAAGGTTGGTTTCGGGTCGAGATCGGACGCTGTGCGCAGA 635  
Db 599 GACCATGGAATCTTCACCATTCGCTTCGCGCTGAGCTAGACACGCTGCGCGGCTC 658  
QY 636 TCTCCCCGAGAACAGCTTCGCGCAGCGGTTTCGATGCGCCAAACATCATCATCATCGCTCGC 695  
Db 659 GCGGGAAGGAGCGCTTCGCGCGGCTTCGAGACGCGCCAGCGAGTTTCAGCTGCTCCG 718  
QY 696 GTTCATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCAGCTCGGTCAGAGGCTTCCT 755  
Db 719 CTAGTGAACGCGTTCGGAAGGCGCAGAGTTCTTCTCGCGTTCGCTCCGAGCGCGGCT 778  
QY 756 AGCGCAGAGCATCAAGCTGTGAGAGTTCACCTACAGCTGATTCGCGCGGAGAGGC 815  
Db 779 CAAGAGCGGCTCAAGGTCGTGCGAGTTCTGTAAGCGCATCCGCGAGAGAGACACA 838  
QY 816 CGAGATCGTCG-----AGGTCCGCGCCAGCGGCAACAGAGAGAGATGAAGACAGACAT 869  
Db 839 GGAGCTTCCGACAGCAAGATGGCGACCGAGGACACTGTCCCTGATTTTCGAGGCAAGACAT 898  
QY 870 CTTCTACCGTTTCATCGAGCTGGCGGAGCGCGCGACGACGCGCGCGCTTCGCGGAGAGA 929  
Db 899 GCTGTCCAGATTTCACACGCGCCAG-----AACGAGTCCGCGGACGGTGGACTA 949  
QY 930 TAAAGCCTCCGGAAGCTGTGTCCTCACTTCGATGTCGCGCGGAGAGAGAGAGAGAGAGAG 989  
Db 950 CAAGTACCTGAGGAGACATCACTACTGAAATAGTATCTGCGCGGCAAGGAGACAGAGCGCGG 1009  
QY 990 GACGCTGTGCTGTTTCAGCAGCATGCGCATGTCCTCCACCGGACGCTGCGCGGAGAGCTCGC 1049  
Db 1010 GCGCTGCGCTGCTGCTTCTTACATGGCTGCAAGCACCCGAGATCCAGAGAGAGATCTG 1069  
QY 1050 CCGGAGCTGTGCGGTTTCGAGCGGAGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1109  
Db 1070 CCAGGAG-----GCCGCAAGGCGACCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1110  
QY 1110 CTGCGCGCGCTGACCGGAG 1169  
Db 1111 -----GCGTGGAGAGAGTTTCGCGCA 1129  
QY 1170 CTTCTTACCTACGACAGCTCGGCAAGCTGCTTACCTCCACGCTGCTGCTCAGCGAGAC 1229  
Db 1130 GAGCTTACCGAGAGGCGCTTGAACAGATGCACTTACTGTCAGCGCGCTGCTGCTGCTGCTG 1189  
QY 1230 GCTCGCTGTATCCCGCGCTTCCTTCAGACCCCAAGGGAGTCTTCGAGGAGAGAGAGAGAGAG 1289  
Db 1190 GCTCAGGCTGTACCGCTCGGTTCTCTGATAAAGCAGTGTCTTCGAGAGAGAGAGAGAGAG 1249  
QY 1290 GCGGAGGAG 1349  
Db 1250 GCTTACGCTTCCAGCGTCAAGGAG 1309

365 GACCTCTCGCGGACGCGCATCTTCAACGTGGACGGCCCAAGTGGCGGACACAGCGGAAG 424  
 430 ACGGCGAGTTTCGAGTTTCGCTCCAAAGAACTGGAGGATTTTCAGGCGCATTTGTGTTCAGA 489  
 425 GTCGCCAGCCAGAGTTTCTCCACGCGGGTGTGCGGACATACAGCACCGGCGGTGTTCGCG 484  
 490 GAGTACTCCCTGAAGCTGTTCGGGTATACGT---AGCCAGGCGATCCAAAGGAGGCAAAATTT 546  
 485 GACACGGCGGCGAGCTCGCGGGCATCGTGGCTCGCGCGCGCGCGCGCGGGGAGAG 544  
 547 GTGACATGACGAGAACTTTATCAGGATGACGTGGATCCATCTGCAAGGTTGGGTTTC 606  
 545 CTAGACATGACGAGAGCTGCTGATGCTTCGACGCTGGAGTCCATCTTCAAGGTCGGGTT 604  
 607 GGGGTGAGATGCGACACGCTGTCGCCAGATCTCCCGAGAGAACAG---TTCCGCGAGCGG 663  
 605 GGGGTACGCTTGGGGCTGCTGTCGGCTCCAGCGAGGAGGGCGGCAATTCGCCAGAGCG 664  
 664 TTCGATGCGGCCAACATCATCATCACGCTGGGGTTTCATCGACCGCTGTGGCGCATCAAG 723  
 665 TTCGACGACGCGAGCAGCAGCAGGTGCTGCGCGCTTCTTGGACCCGTTCTGGAAGGCCAAG 724  
 724 AGGTTCTTCCACGTCGGGTCAGAGCCCTCTCTAGCGCAGAGCATCAAGCTCGTGCAGAG 783  
 725 AGGTTCTTCCACGTCGGGTCAGAGCCCTCTCTAGCGCAGAGCATCAAGCTCGTGCAGAG 784  
 784 TTCACCTACAGCGTGTGATCCGCGGAGGAGGCGGAGATCTGCGAGGTGCGGCGCAGCGCG 843  
 785 TTCGTCACGCGCTGTCGACAGGAGATCTGAGCAGATGAGAAAGAGAC---CNACAGGA 840  
 844 AAACAGGAGAGATGAAGCAGCAGCATCTCTCACGGTTTCATCGAGTCGGCGAGGCGCGG 903  
 841 ATTGTGAGCCCGAGAGATATTACCTCTCCGTAGTCCGTACATGATGGGTGAGTGCAGA 900  
 904 GACGACGCGCGCGCTTCGGGACGATTAAGAGCCTCCGGGAGCTGTGTCTACCTTCATCTCG 963  
 901 CGGACGACGCGGATCACAGAGAGAGATTACAGTTCTGACACCGGTGACATTTCCCTGAC 960  
 964 ATCGCGCGGCGGACACGACGCGCGAGCGCTGTCTGTGTCTACGACATGCGC---ATGT 1021  
 961 GCCAAGAAAGAGGACATCTCTGTCGAGTTCTTGTGGAGGGGAGAGCGATCCCGGCTGC 1020  
 1022 CCCACCGGACGTCGGCGGAGAGCTGCGCGCGAGCTGTGCGGTTTCGAGGCGGAGCGCG 1081  
 1021 TCCGACAAAGAGTACCTTCGGGACATCATCTCACTTGTGATCGCGCGCGGACACG 1080  
 1082 CGCGGAGGAGGCGTCAACGCT-----CGTCTCTGCGCGCGCGCTGACGCGCGAGAC 1134  
 1081 ACCGCGGAGACCTGTCGTGGTTTCTCTACGTGCTCTGACGAGAACAGAGCATCCAGGAC 1140  
 1135 AAG-----GCGTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1161  
 1141 AGGATCCGCGAGAGGTCG 1200  
 1162 TTCGCGGCGCTCTCACTACGACGCTTCGGGAGCTGTCTACCTTCAGCGCTGCGTTC 1221  
 1201 CTGTCGCGGTCCTGACCGGAGGACGCGCATCAGCAAGATGCACTACCTGCAACGCGGCG 1260  
 1222 ACCGAGACGCTTCGCGCTGATACCGCGCTGCTCCAGACCCCAAGGGGATCTCTGGAGGAC 1281  
 1261 ACCGAGACCTTCGCGCTGATACCGCGCTGCTCCAGATGCTGCTGCTGCTGCTGCTGCT 1320  
 1282 GACGTGTCGCGGACGAGAGGTCGAGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1341  
 1321 GACAGCTTCGCGGAGGCTACGCTGTCAAGAAAGGAGACATGTTGACTACCGAGCGCTAC 1380  
 1342 TCGATGGGCGGAGTGGAGTAACTGGGCGCGCGAGCGCGGAGCTTCGCGCGCGGAGCGG 1401  
 1381 CAATGGGCGGAGTGGGTTCTCTGTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 1440  
 1402 TGGATCAAGAGGATGGCGGTTCCGCAACGCTGCGCGCTTCAAGTTTCAAGGCTTCAG 1461  
 1441 TGGTGGACGACGCGCGCTCTTCGTCGCGCGGAGAGCGCGCTTCAAGTTTCAAGGCTTCAG 1500

1350 GCGGATGAGTACAACTCGGGCCCGACGCGCGGAGCTTCGCGCGGAGCGCGGTGATCAA 1409  
 1310 CCGGATGAGTACCTCTGGGCAAGGACGCGGAGGTTCGCGCGGAGCGGTGCTGGA 1369  
 1410 CGAGGATGCGGCTTCGCGCAACGCGTGGCGGTTCAAGTTACGCGGTTCCAGCGGGGCC 1469  
 1370 TCAGGACGAGAGTTCAGCAGGAGAGCGGTTCAATTCACAGGCTTCCAGGCTGGTCC 1429  
 1470 GAGGATGCTGCGGCAAGGATCGGCGTACCTGCGAGATGAAGATGGCGCTCGCCATCT 1529  
 1430 GAGGATGCTGCGGCAAGGATCGGCGTACCTGCGAGATGAAGATTCGCGCGGCTGCT 1489  
 1530 CTTCCGCTTCTACAGCTTCGCGGCTGCGGAGGCGAC-----CGGTCAGATACCGAT 1583  
 1490 CTTCCGCTTCTACAGCTTCGCGGCTGCGGAGGCGACAAAGCGGTGAATTCAGGAC 1549  
 1584 GATGACCATCTCTCCATGCGGACGCGCT 1613  
 1550 CATGATCAGCTCCACATCGACAGGCTCT 1579

RESULT 8  
 US-10-425-114-28465  
 ; Sequence 28465, Application US/10425114  
 ; Publication No. US2004003488A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E.  
 ; APPLICANT: Tabaska, Jack E.  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 28465  
 ; LENGTH: 1947  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB4757-030-F2\_FLI  
 US-10-425-114-28465

Query Match 15.6%; Score 298.2; DB 12; Length 1947;  
 Best Local Similarity 52.8%; Pred. No. 8.9e-70;  
 Matches 817; Conservative 0; Mismatches 678; Indels 52; Gaps 6;

70 CACAAGTACATCGCGCTCTCTGTTGCTCTCTCATGATCTCTGTTCCAGAGTGGAGC 129  
 65 CCCAGGTTCTCCCTGATCTTGTCTCTATCTCTAGTGTCTATCTTGGCGCTCCGGCG 124  
 130 CTGAGGAGCAAGAGCGCGGATGATGCGGATCATGCGGCAACGCTGGAGCAGCTG 189  
 125 CGCGCGCGCGGACAGAACCGGAGTACCGCGCGGTGGCTGCGACCGTCTGCAACAGCTG 184  
 190 AGGACTACACCGGATGACGACTGCTGTGCGGTTACCTGTCTACGCGCAGGACAGTG 249  
 185 TTCAACTTCGCGCGCTGATGAGTACCAACCGCGCTCTCCGCGAGTACCGTACCTTC 244  
 250 ACCGTCGACATCCGCTTCACTTCTTACACCTACATCGCTGACCCCGGTGAATGTCGAGAT 309  
 245 CGCATGCTTACCCGACCTGACGCTTACATCTACCGTGGAGCCCGGAAACGTTGAGTAC 304  
 310 GTCTCAAGACATACTTCAACAAATTAACCAAGGAGATCGTGTACAGATCTTACATGAC 369  
 305 ATCTCAAGGCGCAACTTCGCGCAACTACGCGCAAGGAGGACGCTGACGAGCTGAGCGAG 364  
 370 GTGCTCTCGTGACGCTCTTCAACGCGCGGAGGAGCTGTGGAGGAGCAGAGGAG 429



QY 1462 GCGGGCCGAGGATCTGCTTGGCAAGGACTCGGCGTACTGTCAGATGAAGATGGCGCTG 1521  
Db 1501 GCGGGCCCTCGGCTGCTTGGGAAGGAATTCGCGTACAGGAGATGAAGATCTTCGCG 1560  
QY 1522 GCATCTCTTCGCTTCTACAGCTTCGCGCTGCTGGAGGGGACCC 1568  
Db 1561 GCTGCGCTTCTCTACCTGTTTTCAGGTTTGAGATGGGATGCCAACGC 1607

RESULT 9

US-10-260-238-1236  
; Sequence 1236, Application US/10260238  
; Publication No. US20040016025A1  
; GENERAL INFORMATION:  
; APPLICANT: Budworth, Paul R.  
; APPLICANT: Moughamer, Todd G.  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiyaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
; FILE REFERENCE: 60111-NP  
; CURRENT APPLICATION NUMBER: US/10/260,238  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,448  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 6077  
; SEQ ID NO 1236  
; LENGTH: 1617  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: N region  
; LOCATION: (1326)..(1326)  
; OTHER INFORMATION: n = any nucleotide  
US-10-260-238-1236

Query Match 14.1%; Score 269; DB 15; Length 1617;  
Best Local Similarity 53.2%; Pred No. 6.1e-62;  
Matches 724; Conservative 0; Mismatches 581; Indels 57; Gaps 5;  
QY 282 CATCGCTACCCGGTGAATGTGCGAGCATGTCTCAAGACTAACTTCAACCAATACCCCA 341  
Db 276 CACCTGGACCCGGCAACGTGGCCACATCTTCACTTCACTTCACTTCACTTCACTTCACT 335  
QY 342 GGAATCTGTACAGATCTTACATGACGTGCTCTCGTGAAGGATCTTCAACGCCGA 401  
Db 336 GGGCGACGTGTTCG---CCGACATGTTCACTTCTCGCGGGCGGATCTTCAACGCCGA 392  
QY 402 CGGCGAGCTGTGGAGAACAGAGGAGACGGCGAGTTTCGAGTTTCGCTTCAAGAACCT 461  
Db 393 CGGCGAGCGGTGGCGGGCGGACGGGACAGGCCAGATGCTCATGACACCGCGGTT 452  
QY 462 GAGGATTTTCAGCGCATTTGTTCAGAGAGTACTCCCTGAAG---CTGTGGGTATATCT 518  
Db 453 CCGCGCTTCTGTGGCGGCTCCAGCTTCGACAGGTGGAGAGGGGCTCTGTGCTTCTCT 512  
QY 519 GAGCCAGGATCCAGAGGAGGCAAAATTTGTGACATGAGGAACTTTATGAGGATGAC 578  
Db 513 CGGCGACGTGGCGGACCGGAGACGTGCGACCTGCAGGACGTGTTCAOGCGGTGGTC 572  
QY 579 GCTGGACTCATCTGCAAGGTTTGGTTTGGGTCGAGATCGGCAACGCTGTTCGCGAGATCT 638

Db 573 GCTCGACGCGAGCTGCCACTGGTGTTCGCGCTGCAACCCCGGTGCTTGGACATCGGCT 632  
QY 639 CCCGAGAACAGCTTCGCGCAGCGCTTCGATCGCGCAACATCATCATCATCGCTCGGCTT 698  
Db 633 GCGGAGGTTCCTTCGCGCGCGGATGAGACGCTGTCGCGACCATCTTCTTCGCGCA 692  
QY 699 CATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTTCAG-----AGGC 749  
Db 693 CACCATGCCAGTGTGTTCTGGAAGACGATGAGATGGCTAGGATTTGGGCAACAGAGAG 752  
QY 750 CTTCTTAGCGCAGAGCATCAAGCTCTGTGAGAGGTTCACTACAGCGTGTATCGCGCGGAG 809  
Db 753 GAACGCGCGCGGAGCGGACGCTCGAGAGCTTCGTTCGCGTCGCGCATCGGCAACACAG 812  
QY 810 GAAGGCGGAGATCGTCGAGTTCGCGGCGCAGCGGCAAAACAGAGGAGATGAAGACACGAT 869  
Db 813 AGCGGATGACGAGAGCAAGTTCAGACAAGGCGGCGGAGGCGGCGGCAAGCAATCGGT 872  
QY 870 CTTGTCACTGTTTCATCGAGCTGGGCGAGGCGCGGCGAGCGAAGGCGGCGGCTTCGGGACGA 929  
Db 873 CGCGGAGTTGCTTTCCTTCCTTCCTTCGACGACGAGATCGCGGCGAGCGCGCGCA 932  
QY 930 TAAGAGCCTCCGCGAGCTGTGTCTCACTTCGTGATCGCGGCGGCGGACACGCGGCGAC 989  
Db 933 CGTGTACATCAGGAGCATGCGCGATGAACCTCTCTGTCGCGCGCGGCGGACACCGAGCTC 992  
QY 990 GACGCTGTCTGTGTTCAACGACATGGCATGTCCACCCGAGAGCTGGCGCGGAGAGCTGCG 1049  
Db 993 GCGCTGTCTGTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1052  
QY 1050 CCGCGAGCTGTGCGGTTTCGAGGCGGAGCGCGGCGGCGGAGGCGCTCAGCTCTGCT 1109  
Db 1053 GCAGGAGCTGGGCGCATCG-----CTTCCGCAAGCGCGAGCTGCGAGCAGGCGG 1103  
QY 1110 CTGCGGCGGCGCTGACCGCGAGCAAGGCGTTTCGCGCGCGCTGCGCGAGTTCCGCGG 1169  
Db 1104 GCTCTTCCCGGCAACCGCGGCTGTCGAGCGGAGCGAGTGAGG-----1155  
QY 1170 CTTCTCACCTACGACAGCTCGGCAAGCTGTCTACCTCCAGCTTCGCTTACCGAGAC 1229  
Db 1156 -----AACCTCTGTACTTGCACGCGGCTTTTCGAGGC 1190  
QY 1230 GCTCCGCTGTACCCCGCTTCCTCAGGACCCCAAGGGGATCTTGGAGGACGAGTCT 1289  
Db 1191 CATGAGCTGTACCCCGCGGTGCTTTGGAGCACAAGGCGCGGTTCGCGCAGCGCTGCT 1250  
QY 1290 CCGCGACGAGGAGTGTGAGGCGCGGCGGATGCTGACGTACGTCCTTCTTCTTCTTCTT 1349  
Db 1251 CCGCAGCGGCGACGAGGTGATGGCGCGGACAAGGTGCTGTGTTTCTACTACTCATGGG 1310  
QY 1350 GCGGATGAGTACAACTGGGCGCGCGAGCTTCGCGCGGAGCTTCGCGCGGAGCGGTGATCAA 1409  
Db 1311 GAGGATGAAGCGGGTNTGGGCAAGGACTGCGAGGAGTTTCAAGCGGAGCGGTGATCAC 1370  
QY 1410 CGAGATCGCGGTTTCCCAACGCTGCTGCTTCAAGTTTCAAGGTTTCAAGGCGGCGGCG 1469  
Db 1371 CGAGGACGCAAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1430  
QY 1470 GAGGATCTGCTTGGCAGAGGCTGCGGCTTCTGAGATGAAGATGGGCTTGGCGCATCT 1529  
Db 1431 GAGGACATGCTTGGCAAGGAGATGGCGCTCTGTGAGATGAAGTCAAGGCGGCGGCTAT 1490  
QY 1530 CTTCCGCTTCTACAGCTTCGCGCTGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1589  
Db 1491 GCGGTGGAATTTGCGCTGGAGGTGGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1550  
QY 1590 CATCTCTTCCATGGCGGCGGCTCAAGGTTCGCGCTCTCTCTCTCTCTCTCTCTCTCT 1631  
Db 1551 CATCTCCACATGAAGATGGGCTCTTGTGAAGGTCAAGAG 1592





; NUMBER OF SEQ ID NOS: 6077		
; SEQ ID NO 907		
; LENGTH: 1638		
; TYPE: DNA		
; ORGANISM: Oryza sativa		
US-10-260-238-907		
Query Match		
Best Local Similarity 12.0%; Score 228; DB 15; Length 1638;		
Matches 735; Conservative 0; Mismatches 500; Indels 117; Gaps 9;		
Qy	290	ACCCGGTGAATGTGCGAGATGCTCTCAAGTAACTTACCAATTCACCAAGGGAATCG 349
Db	293	ACCCGGGAACTTTGAGACATCTCTGGCGGCGCTTCGAACTACCCCAAGGGCCCCA 352
Qy	350	TGTACAGATCTACATGACAGCTGCTCTCGTGTGACGCGCATCTTCAACGCCGACGGCGAGC 409
Db	353	TGTGGCAGGATCGTTCCAGACCTCTCTGGCCAGGGGATCTTCAACTCCGACGGCGAGA 412
Qy	410	TGTGGAGGAAGCAGAGGAAGCGGAGTTTCGAGTTTCGCTCTCAAGAACTGAGGAT 469
Db	413	CGTGGCTGATCGACGAAAGACGGCAGCCCTCGAGTTTCAACCCCGGACGTTGCGCCAGG 472
Qy	470	TCAGGCCATGTGTTTCAGAGTACTCTCCCTGAA---GCTGTGCGGTATCTAGAGCCAGG 526
Db	473	CCATGGACAGCTGGGCCAACCGCTCCATCAAGTACCGCTGTGGGTTATCTTCGACGACC 532
Qy	527	CATCCAGGCGGAAAGTTTGTGGACATCGAGGAATTTATCATGAGGATGACGTGGACT 586
Db	533	ACTGCACGGCGCCGACAGCGTGGACCTCCAGGACCTCTCTGCGCTCACTTCGATA 592
Qy	587	CCATCTGCAAGTTGGGTTGGGGTTCGAGATCGGCACGCTGTGCGCAGATCTCCCGGAGA 646
Db	593	ACATCTGGGGCTCACTTTCGCAAGAGCCACGAGACGCTGTCCCGCGCTTGGCGGAGA 652
Qy	647	ACAGCTTCGCGCAGCGTTTCGATGCGGCCCAACATCATCATCATCGTGGGTTCATCGACC 706
Db	653	ACCCCTTCGCCACCGCTTTTGACGAGGCCACCGAGGCGACGATGACAGGTTCTGTTC 712
Qy	707	CGC-----TGTGGCGCATCAAGAGTTCTTCCAGCTCGGFTCAGAGGCCCTCTTAGCGC 760
Db	713	CGAGCTTGTGTGGCGCATCAAGAGGCTTCGTTGTCGGCAGCGAGCGGAGCTTCCGCG 772
Qy	761	AGAGCATCAAGCTGTGAGCAGATTCACTACAGCGTGTATCCGCCGGAGGAGGCCGAGA 820
Db	773	ACAGCCTCGCCGTGTGGA-----CCGTCACATGACGGAGA 808
Qy	821	TCGTCGAGTTCGGCGCAGCGGCAACAGGAGAAGATGAAGCAGCATCTCTGTCACTG 880
Db	809	CCATCGGCACG-----CAAGCGACGCCGTGCGACGACCTGTCTCCCGT 856
Qy	881	TCATCGAGCTGGCGAGCGCGCAGCAGCGCGGGCGGCTTCGGGGACGATAGAGCTCC 940
Db	857	TCATGAAGAAGCGCA-----CAGCAAGGCAAGCGTTCCCGGAGGACGTG---CTCC 907
Qy	941	GGACGTGTGTCTCACTTCGTGATCGCGGGGGGACACGACGCGGACGCTGCTCT 1000
Db	908	AGTGGATCGCTCACTTCGTGCTCGCGGGCGGACACGCTGTCTGCTGCGGCTCAGCT 967
Qy	1001	GGTTTCAGCACATGGCCATGTCTCCACCCGCGAGCTGGCGCGAGAGTTCGCGCGAGCTGT 1060
Db	968	GGTTCTTCTGATCTCATGACAGAGGGGACGCTGGAGCGCAAGTGTCTCTCGAT-- 1025
Qy	1061	GCAGCTTCGAGGCGGAGCGCGCGCGAGAGGGCGTCAAGCTGCTGCTGCTGCGGGCGG 1120
Db	1026	-----CGCTTCGTTCTCAGGAGACGCGCGGGGACGACACACGCGGAGG 1068
Qy	1121	CTGACGCCGACGACAAAGCGCTTCGCCCGCGCTGGCGGAGTTGCGGGCTCTCACT 1180
Db	1069	TGAGCG-----GAGGAGCCTCTAACT 1090
Qy	1181	ACGACAGCCTCGGCAAGCTGTCTTACCTCCAGCCTGCTGCTACCGAGACGCTCCGCGTGT 1240

Db	1091	TCGACGAGCTGGAGCGACTGTGTACTCAAGCGCGCATTTGACGGAGACGCTCGGCTGT 1150
Qy	1241	ACCCGCGCTCTCTCAGGACCCCAAGGGGATCTCTGAGGACGACGCTGCTGCCCGACGGGA 1300
Db	1151	ACCCGCTGTGCGCGCAGGATTCCAAGTACGTCTGCGGACGACGCTGCTGCCCGACGGCA 1210
Qy	1301	CGAAGGTGAGGGCGCGCGGATGGTGACGTACGTGCCCTACTCGATGGGCGGATGAGT 1360
Db	1211	CAGTGTGCGCGCGCTCGGCGATCACCTACTTCCATCTACTCAGTGGGAAGATGAGAGA 1270
Qy	1361	ACAACTGGGGCCCGGACCGGCGAGCTTCCGCGCGGAGCGGTGGATCAACGAGGATGGCG 1420
Db	1271	GCATATGGGGGAAGGACTGCGCAGAGTTCCGGCGGAGCGGTGCTATCGCGCGACGGCA 1330
Qy	1421	CGTTCGCCAA-----CGCTCGCGCTTCAAGTTTCAAGGCGGGGCGCGGAGGA 1474
Db	1331	GCGCTTTCGAGCCTGTCAAGGACGCTACCGCTTCTGTTGGGTTCAACGGCGGCGCGGAA 1390
Qy	1475	TCGCTCTGGGCAAGGACTCGGCGTACCTGCAGATGAAGATGGGCGTGGCCATCTCTTCC 1534
Db	1391	CGTGCCTCGGCAAGGACCTGCGCTTACCTGCAGATGAATCCATCGCTCGCGCTGCTGTC 1450
Qy	1535	GTTTCTACAGCTTCGCGCTCTGGAGGGGACCGGTGCGAGTACCGCATGATGACCATCC 1594
Db	1451	TGCGCAACTCGGTGAGCTCGTGCCTGGGCGCACAAAGGTGGAGCAGAGATGTGCTGACGC 1510
Qy	1595	TCCTCATGGCGCAGCGGCTCAAGGTCCGCGTC 1626
Db	1511	TTTTCATGAAGATGGACTCGGAGTGCACGTC 1542

RESULT 12

US-10-260-238-422

; Sequence 422, Application US/10260238

; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.

; APPLICANT: Moughamer, Todd G.

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Provart, Nicholas

; APPLICANT: Rieke, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION

; FILE REFERENCE: 60111-NP

; CURRENT APPLICATION NUMBER: US/10/260,238

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04

; NUMBER OF SEQ ID NOS: 6077

; SEQ ID NO 422

; LENGTH: 1527

; TYPE: DNA

; ORGANISM: Oryza sativa

US-10-260-238-422

Query Match 11.9%; Score 226.4; DB 15; Length 1527;

Best Local Similarity 54.2%; Pred. No. 1.7e-50;

Matches 678; Conservative 0; Mismatches 476; Indels 96; Gaps 7;

Qy	266	TCACITCTCTACACCTACATCGTGTGACCCCGGTGAATGTTCGAGCATGTCTCAAGACTAACT 325
Db	239	TCACGGCGGGGTCTATCACCGCAACCCCGCCANTGTCTGAGTATACGCTGAGACTAACT 298
Qy	326	TCACCAATTAACCCCAAGGAATCGTGTACAGATCTCTACATGACGCTGCTCTCGGTGACG 385

Db	299	TTGTTAACTACCCAAAGGCGAGCTCGCCGTGTCATCTCGTCAACTTCTTCGGCCATG	358
Qy	386	GCATCTCAACGCGCGAGCTGTGGAGAAAGACAGAGAAAGACGCGAGTTTCAGT	445
Db	359	GCATCTTCACTCGACCGCGAGCGTGCAGTGGCAGCGAAGCCGCCAGCTACGAGT	418
Qy	446	TCGCTCTCAAGAACCTGAGGATTC---AGCGCAATTGTGTTCAGAGACTCCCTTGA	502
Db	419	TCAACAAGCGCTCGCTGAGGAATTCGTGGTGGACACCGTCAGGTCGAGTTCGTCAGA	478
Qy	503	AGCTGTGGGTATCTAGCCAGGATCCAGGCAAGCAAGTGTGACATGACGAGAAC	562
Db	479	GGCTCTCGCTGTGGAGCGGGCGGAGCGCGCGAGCGCTGACGTGCGAGGAG	538
Qy	563	TTTACATGAGGATGACGTGGAATCCATCTGCAAGGTTGGGTCGGGGTCGAGATCGGCA	622
Db	539	TGCTGAGCGCTTCGGTTCGACACATCTGCATGTGCGCTTCGACGAGGA-----	590
Qy	623	CGCTGTGCGCAGATCTCCCGAGAACAGTTTCGCGAGCGGTTTCGATGCGCGCAATCA	682
Db	591	-----CCGCGCTGCTCGCGAGGACAGCATGGC---CTCGCCCGCAGAGCGCAAGTTCA	643
Qy	683	TCATCAGCTGGGTTTCATCGACCCGCTGTGGCGATCAAGAGTTCTTCCACGTCGGT	742
Db	644	TGCGCGCTTCACTGACGCGCAAGATGCGGTGATGACCGGTTCACTGTCGCGTCAAGT	703
Qy	743	CAGAGCCCTCTAGCGCAGAGCATCAAGCTGTGGACGAGTTTCACCTACAGCTGATCC	802
Db	704	CGCGTGGCGCTTCAAGAGCGCTTCAACATGAGCCCGAGAGGAGGATGCGGAGGCG	763
Qy	803	GCGGAGGAAGCGCGAGATGCTCGAGGTCGGGCGAGCGCGCAACAGGAGAGATGAAC	862
Db	764	TGCGCACGATCCACGCTTCGCGAG---CGGATCTGTCGGGAGCGACGCGCGGCTTCG	820
Qy	863	ACGACATCTGTACAGTTTCATCGAGCTGGCGAGCGCGCGACGACGCGCGGCTTCG	922
Db	821	AGCGCGGCTTGGCGCGAGTGACATCTCTGTGCGCTTCGCGCGAGGCGGATCA	880
Qy	923	GGACAGTAAAGCCTCCGGGAGTGTGTCTCAACTTCGTGATGCGCGCGGCGGACGA	982
Db	881	GCACG---AGAGCTCCGTGACGTGGTGAACCACTTCTCATCGCGGCGCGACACGA	937
Qy	983	CGCGCAGCAGCTGTCTGTGTTTCAGCAATGCGCATGTCCTCCACCGAGCTGCGCGA	1042
Db	938	CGTCAAGCGGTTACCTGTTCTTCTGGCTACTGTTCGCGCGCGCGCGAGTGGAGCA	997
Qy	1043	AGCTGCGCGGAGCTGTGCGCTTCGAGCGGAGCGCGCGAGGCGGCGTCAAGC	1102
Db	998	AGATCGTACGAGATCCACGCA-----	1020
Qy	1103	TCGTGCTCTCGGCGCGCTGAGCCGACGACAGCGGTTTCGCGCGCGCGTGGCGAGT	1162
Db	1021	---GTGCGGCGCGCTTCGCTGGGACAGCGACCGACGCT-----	1057
Qy	1163	TCGCGGCGCTCTCACTACGACAGCTTCGCAAGCTGTCTTACTCCAGCGCTGCTCA	1222
Db	1058	-----TCAACTTGGACGAGCTGGGTGACATGCAATACCTCCATCGCGCCATCA	1105
Qy	1223	CCGAGAGCTCCGCTGTACCCCGCGCTCCCTCAGGACCCCAAGGGGATCTGAGGAGC	1282
Db	1106	CCGAGTCCATCGGCTGTACCCACCGTGGSCATGAGCTCGACAGCTGCAAGAGGAGC	1165
Qy	1283	ACGTGCTCCCGGACGGGACGAAGGTGAGGGCGCGGAGAGGTGACGTAGTGGCTTACT	1342
Db	1166	ACTTCTTCCGCGACGGCAGCTGTGTGGGAAAGGTTGTTGTCTTACAGCGCTACG	1225
Qy	1343	CGATGGGCGGATGAGTACAACTTGGGCGCCGAGCGCGCGAGCTTCGCGCGGAGCGGT	1402
Db	1226	CGATGGCAAGGTGAGGACATCTGGGCGCGGAGCTGTGAGGAGTTCAAGGCGGAGT	1285
Qy	1403	GGATCAAGGAGTGGCGGCTTCGCAACGAGTTCGCGCTTCAAGTTCAAGCGCTTCAGG	1462

Db	1286	GGCTAGACGAGCGCGCGCTTCGCGCGGAGAGCCGTTTCAAGTACCCGGTTTCCACG	1345
Qy	1463	CGGGCCGAGATCTGCTGGCAAGACTCGGGTACCTGCAGATGAAG	1512
Db	1346	CGGGCCGAGATGTGCTTCGCGCAAGGAGATGGCTTACATACAGATGAAG	1395
RESULT 13			
US-10-425-114-1524			
; Sequence 1524, Application US/10425114			
; Publication No. US2004003488A1			
; GENERAL INFORMATION:			
; APPLICANT: Liu, Jingdong			
; APPLICANT: Zhou, Yihua			
; APPLICANT: Kovalic, David K.			
; APPLICANT: Screen, Steven E			
; APPLICANT: Tabaska, Jack E			
; APPLICANT: Cao, Yongwei			
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With			
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement			
; FILE REFERENCE: 38-21(53313)B			
; CURRENT APPLICATION NUMBER: US/10/425,114			
; NUMBER OF SEQ ID NOS: 73128			
; SEQ ID NO 1524			
; LENGTH: 1680			
; TYPE: DNA			
; ORGANISM: Zea mays			
; FEATURE:			
; OTHER INFORMATION: Clone ID: 700156359_FLI			
US-10-425-114-1524			
Query Match			
Best Local Similarity 11.4%; Score 217.2; DB 12; Length 1680;			
Matches 705; Conservative 0; Mismatches 543; Indels 102; Gaps 6;			
Qy	286	GCTGACCCCGGTGAATGTTCGAGCATGCTCTCAAGATAAATTACCCCAAGGGA	345
Db	342	GCCGACCCCGGCAACGTGCGCACGTTTCACTCCAACTTCGGCAACTTACCCCAAGGC	401
Qy	346	ATCTGTATCAGATCTTACATGACGCTCTCTCGGTGACGGCATCTTCAACCGCCACGC	405
Db	402	CAGGAGTTTCGCGCGCTGTTCGACGTG---CTCGCGACGGCATCTTCAACCGCCACGC	458
Qy	406	GAGCTGTGAGGAGACGAGG---AAGACGGCGAGTTTCGAGTTCGCTTCCAGAACCTG	462
Db	459	GAGTGTGGCGGTTTCAGAGCGGACGCGCTCTCTCGGACGCGGTTCCCG	518
Qy	463	AGGGATTTTCAGCGCCATTGTTCAGAGAGTACTCCCTGAAAGTGTTCGGGTATCTGAGC	522
Db	519	GCGGGGTTCGCGCCACGACCGCGCAAGCTTCGCGACGCTCTCTGTCGCTCCTCGAC	578
Qy	523	CAGGATTCAGGACGCAAGTTGTGGACATGACGAGAACTTTTACATGAGGATGAGCTG	582
Db	579	GGCATTCGCGCGCTTCGCGCGGCTGAGGACCTTTCGAGGACGTTGTCGCGCTCAGCTTC	638
Qy	583	GACTTCATCTGAAAGTTCGGTTTCGGGTTCGAGATCGGACGCTGTCGCGAGATCTCCCG	642
Db	639	GATCTCACCGCATGTTTCGTTTCGCAATCGACCCCGGCTGCTCCGCTGCTGACTTCCCG	698
Qy	643	GAGAACAGCTTCGCGCAGGCTTCGATGCGCCCAACATCATCATCAGCTGCGGTTTATC	702
Db	699	TACGTCCCGTTTCGCGCGCCATGGACACCGTTCGAGGAGTCTCTGTTTACCGGACGTC	758
Qy	703	GACCCGCTGTCGCGATCAAGAGTTTCTCCAGTCTGGGTTCAGAGCCCTTCTAGCGCAG	762
Db	759	GCGCCGCTGCGGTGCTGCTCCAGAGTACTCAGGAC-----TGGGCAC	806
Qy	763	AGCATCAAGCTTCGAGGAGTTTCACTACAGCGGTGATTCGCGGAGGAGGCGCGAGATC	822
Db	807	AGCAAGAGATGTGGAAG-----CTCGACGGGTGCTCGACGCTCCATCGCGGAGTTC	860
Qy	823	GTGAGGTTCGCGGCGCGCAACAGGAGAGATGAACGACATCTCTGTCAAGGTTTC	882



Db 1144 -----TCCGCGAGATGGCTACGTGCA CGCGCCAT 1174  
Qy 1221 CACCGAGAGCTCCGCTGTATCCCGCGTCCCTCAGAGACCCCAAGGGGATCCTGGAGGA 1280  
Db 1175 CACGAGAGTGTATCGCGCTGTATCCCGCGTCCGCTGAACTCGCTGTGGCGGAGACGGA 1234  
Qy 1281 CGACGTGTCTCCCGGACGGGACGAAGAGTGTAGGGCCCGCGGGATGTGTACGTACGTGCCCTTA 1340  
Db 1235 CGACGTCTCTCCCGGACGGGACGAGCGGTGTGTCAGGGTGTTCGTGTGGGTACAATTCTATA 1294  
Qy 1341 CTCGATGTGGGCGGATGTAGTA CAACTGGGGCCCGGAGCGGCGGAGCTTCGGCCCGGAGG 1400  
Db 1295 TGGGATGGGCGGATGTAGTTCGCTGTGGGGCGACGACGCGCGGGAATACCGGCCGAGCG 1354  
Qy 1401 GTGGATCAA--CGAGGATGGGCGGTTCGCAACGCGTCCGCGTTCAAGTTTCAACGCGCTT 1457  
Db 1355 GTGGCTGAACCCGAGAGCGGAAGCTTCGGCGCGGACAGCCCTTTCGGTTCGTGGCAT 1414  
Qy 1458 CCAGCGGGGCGGAGGATCTGCTTGGGCAAGGACTCGGCGTACTCGGCTACCTGTAGATGAAGTGC 1517  
Db 1415 TCAGCGGGGCGGAGCTTGTCTGGGCAAGGAATGGCGTATCCAGATGAAGTCCAT 1474  
Qy 1518 GCTGGCATCTCTTCCGCTTCTACAGCTTCGGCTGTGGGAGGACCCCGTGTAGTA 1577  
Db 1475 CGTGGCTGTCTGTGGAGAGCTCGACGTGGCGGTGTAGCGGCGGTACCGGCCCGGCA 1534  
Qy 1578 CCGCATGATGACCATCTCTCATGCGGACGCGGCTCAAGGTCCGCGT 1625  
Db 1535 GGTGGCTGTCTAAGCTGGGATGCGGACGCGGCTACCTGTGACGT 1582

## RESULT 15

US-10-425-114-26392  
; Sequence 26392, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-26  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 26392  
; LENGTH: 1761  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB4171-038-F6\_FLI  
; US-10-425-114-26392

Query Match 10.4%; Score 198; DB 12; Length 1761;  
Best Local Similarity 52.1%; Pred. No. 8.1e-43;  
Matches 730; Conservative 0; Mismatches 525; Indels 147; Gaps 7;

Qy 175 ACGTGGAGAGCTGAGAACTACACGGGATGACGAGCTGGCTTGTGGGTACCTGTCA 234  
Db 240 ACGGGTGTGGGCGCGGACTTACCAACCTCGGCGACTGGTACGCGCACTGCTCCGG 299  
Qy 235 CGGACAGGACAGTACGCTGCATGCGCTTCTTCTACACTACATCGCTGACCG 294  
Db 300 GAGTTCGCCCGGACCATCCAGTGCA CGTGTCTGGCTGCACCTGTACCGCAACCCG 359  
Qy 295 GTGAATGTGAGCATGTCTCAAGACTAATCTACCAATTAACCCCAAGGGAATCGTGTAC 354  
Db 360 GCCAAGCTCGAGTACATGCTGAGGACCACTTCGACAACTTCCCAAGGGGAAGGCTTC 419

Qy 355 AGATCTTACATGAGCTGTCTCGGTGACGGCATCTTCAACGCGGACGCGAGCTGTGG 414  
Db 420 GCGGCGTGTCTGGGGACCTTCTCGGCGGCGCATCTTCAACGTGGAGCGCCACGCTGG 479  
Qy 415 AGGAAGCAGAGGAAGACGGGAGTTTCGAGTTCCCTTCCAAGAACCTTGAGGGAATTC--- 471  
Db 480 CGCCACACAGCGAAGATGGCCAGCTCGAGCTCGGCAGCGTGGCGCTCCGCTCTACGCG 539  
Qy 472 AGCGCATTTGTTCAGAGAGTACTCCCTGAGCTGTGGGTATACCTGAGCCAGGATCC 531  
Db 540 TTCGGATTCGTGGCGCAGGAGTGGAGTCCCGCTCTGTGCGGTGCTTCGCCGCCGCCGCC 599  
Qy 532 AAGCAGCAGAAAGTTGTGGACATCAGGAATCTTACATGAGATGAGCTGGAATCCATC 591  
Db 600 GACGTGGCGGCTGTGGACCTTGAGGACGTGTTCGCGCTTCGCGTTCGACACCATC 659  
Qy 592 TGAAGGTTGGGTTTCGGGTCGAGATCGGACCGTGTTCGCGCAGATCTTCCCGAGAACG 651  
Db 660 TGAAGATCTCTTTCGGCTTGAACCCCGGATGCTTGGAGCGGAGATGTCGCGTGTTCGAGG 719  
Qy 652 TTCGGCGAGGCTTCGATGCGCGCAACATCATCATCATCGCTCGGTTTCATCGACCG--- 708  
Db 720 CTCGCGACGCTTCGACGTCCGACGCGCTGTGCGCATGCGCGGCGCGCGCGCTCG 779  
Qy 709 -----CTGTGGCGCATCAAGAGGTTCTTCCAGCTCGGCTCAGAGGCGCTCTCTACGCGAG 762  
Db 780 CCGTAGTGTGAGGGCCAGCGCTCTGAAAGTCTGGTCCGAGAGGAGCTCCGCGGAG 839  
Qy 763 AGCATCAAGCTGTGGACGAGTTACCTACAGCTGATCCGCGGAGGAGGAGCGCGATC 822  
Db 840 GCCATCGGCTGTGACGAGCTCGCGCGCGCTGTATCCGCGGAGCGCGGAGCTGGGA 899  
Qy 823 GTGAGGTCGGGCGGCGGCAACAGGAGAGATGAGACGACATCTGTACGCTTC 882  
Db 900 CTCGTGGCGAG-----CCACGACCTCTCTGTCCGCTTC 932  
Qy 883 ATCGAGCTGGGCGGCGGCGGCGGACGACGCGGCGGCTTCGGGACGATAAGAGCTTCGG 942  
Db 933 ATGGCTTCGGCGGCGCGGCGA-----CGGCGACGACAGTACCTCGC 977  
Qy 943 GACGTGGTCTCAATTGTGTATTCGCGGGCGGAGACGACGCGGACGAGCTGTGTGG 1002  
Db 978 GACATCGTGTGAGCTTCTGTGCGCGGCGGACACGCTGTCTCTCGCGCTGAC----- 1033  
Qy 1003 TTCACGACATGGCATGTCCACCGGACGTGGCGGAGAGCTGGCGCGGAGCTGTGC 1062  
Db 1034 -----CAGCTGTTC 1043  
Qy 1063 GCGTTCGAGGCGGAGCGCGCGGCGGAGAGGCGTCAACGCTCTGTGTGTGGCGCGCT 1122  
Db 1044 ATGCTCTTCCCAAGAACCCCGGCGGTGGCGCGCATCCGCGG----- 1088  
Qy 1123 GACGCGACGACAAAGCGTTTCGCGCGGCGGCGAGTCCGCGCGCTCTACCTAC 1182  
Db 1089 -----GAGGCGGCGGCGGCGAGTCCGCGCGCTCTACCTAC 1127  
Qy 1183 GACAGCTTCGCAAGCTGTCTACCTCCACGCTGTGTCACCGAGACGCTCCGCGCTGTAC 1242  
Db 1128 GAGCACTCAAGGCGCTGACATACACCCAGCGGCTGTGTACGAGAACATCGGCTGTTC 1187  
Qy 1243 CCGCGCTCTCTCAGGACCCCAAGGGATCTTCGAGGAGCAGCGTGTGTGGCGGAGC 1302  
Db 1188 CGCGCGTGTGAGTTTGACTTCAAGTTCTCGCGGGGCGGAGCTGTCTCCCGACGCGACC 1247  
Qy 1303 AAGTGTGAGGCGCGGCGGATGTGTGAGTGTGCTTCTCTGATGGGCGGATGGAGTAC 1362  
Db 1248 TACGTGGCGGCGGCGGCGGCTGTACCTACCCCTTACGCAATGGCGCGGATCCGCGC 1307  
Qy 1363 AACTGGGCGCGGACCGCGGCGGAGTTCGCGCGGAGCGGTGGATCAACGAGGATGGCG-- 1420  
Db 1308 ATCTGGGCGGCGGACCAACGCGCGCTTCCGCGCGGCTTCCGCGCGCTACCGGAGCGCGCC 1367  
Qy 1421 -CGTTCGCAACGCTCGCGCTTCAAGTTTCACGCGCTTCCAGGCGGCGGCGGAGATCTGC 1479

Db	1368	TCGTTTCCTCCAGAGAGCCTGTATAGGTACCCGGTGTTCAGGCGGGCCCTCCGCTCTGC	1427
QY	1480	CTGGGCAAGGACTTCGGCGTACCTGCAGATGAAGATGGCGCTGGCCATCCTCTTCCGCTTC	1539
Db	1428	CTCGGCAAGGAGCTCGCCGTCACCGAGATGAAGTGGTCTGCTGGCCGCTGGTGAGGGCG	1487
QY	1540	TACAGCTTCGGGTCTGGAGG	1561
Db	1488	TTCGAGTCGAGGTGGTCGGGG	1509

Search completed: March 6, 2004, 04:20:59  
Job time : 707 secs





DR	WPI; 2002-471258/50.
DR	P-PSDB; ABH08037.
XX	Novel nucleic acid sequence comprising SBMu200 gene, useful for producing transgenic plants and mediating male fertility in plants.
PT	
PT	
XX	Claim 3; Fig 4; 60pp; English.
XX	The invention relates to a SBMu200 gene, useful for imparting male fertility in a plant e.g. maize. An isolated nucleotide sequence comprising the SBMu200 gene is useful for producing hybrid seeds and for providing heritable externally controllable male sterility in a plant.
CC	The present sequence represents the maize SBMu200 cDNA
XX	
SQ	Sequence 1906 BP; 398 A; 561 C; 586 G; 361 T; 0 U; 0 Other;
	Query Match 100.0%; Score 1906; DB 6; Length 1906;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 1906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GAATTCGGACAGGGAAGCTCACCTCAGCGCGGACGCCATCGCCATTCTTCCCACCTA 60
DB	
QY	1 GAATTCGGACAGGGAAGCTCACCTCAGCGCGGACGCCATCGCCATTCTTCCCACCTA 60
DB	
QY	61 GCAGGGCCTCACAAGTACATCGGCTTCCTTGTTGTTCTCTCATGATCCTGTTCCAG 120
DB	
QY	61 GCAGGGCCTCACAAGTACATCGGCTTCCTTGTTGTTCTCTCATGATCCTGTTCCAG 120
DB	
QY	121 AGGTGAGCTGAGGAAGCAGAAGGCCCGAGATCATGCCAGTCATCGGCGCAACGGTG 180
DB	
QY	121 AGGTGAGCTGAGGAAGCAGAAGGCCCGAGATCATGCCAGTCATCGGCGCAACGGTG 180
DB	
QY	181 GAGCAGCTGAGGAATACCACCGGATGCAGACTGGCTTTGTCGGGTA CTGTACGGCAC 240
DB	
QY	181 GAGCAGCTGAGGAATACCACCGGATGCAGACTGGCTTTGTCGGGTA CTGTACGGCAC 240
DB	
QY	241 AGGACAGTGAACGTGAGCATGCGTTCACTTCTACACCTACATCGCTGACCGCGTGAAT 300
DB	
QY	241 AGGACAGTGAACGTGAGCATGCGTTCACTTCTACACCTACATCGCTGACCGCGTGAAT 300
DB	
QY	301 GTCCAGCATCTCCTCAAGACTAACTTCAACCAATTACCCCAAGGAATCGTGTACAGATCC 360
DB	
QY	301 GTCCAGCATCTCCTCAAGACTAACTTCAACCAATTACCCCAAGGAATCGTGTACAGATCC 360
DB	
QY	361 TACATGACAGTGTCTTCGTGACGCGCATCTTCAAAGCGGACGCGAGCTGTGGAGGAAG 420
DB	
QY	361 TACATGACAGTGTCTTCGTGACGCGCATCTTCAAAGCGGACGCGAGCTGTGGAGGAAG 420
DB	
QY	421 CAGAGGAGAGCGGAGTTTCGAGTTTCGCTTCCGCTCCAAGAACCTGAGGATTTTCAAGCGCAT 480
DB	
QY	421 CAGAGGAGAGCGGAGTTTCGAGTTTCGCTTCCGCTCCAAGAACCTGAGGATTTTCAAGCGCAT 480
DB	
QY	481 GTGTTTCAGAGAGTAGTCTCCCTGAAGCTGTCCGGTATATGAGCCAGCGCATCAAAGGAGGC 540
DB	
QY	481 GTGTTTCAGAGAGTAGTCTCCCTGAAGCTGTCCGGTATATGAGCCAGCGCATCAAAGGAGGC 540
DB	
QY	541 AAAGTTGTGGACATGCAGGAACCTTACATGAGGATGACGCTGCACTTCCATCTGCAAGTTT 600
DB	
QY	541 AAAGTTGTGGACATGCAGGAACCTTACATGAGGATGACGCTGCACTTCCATCTGCAAGTTT 600
DB	
QY	601 GGTTTTCGGGTTCAGATCGGCACGCTGTGCGCAGATCTCCCGAGAACAGCTTTCGCGCAG 660
DB	
QY	601 GGTTTTCGGGTTCAGATCGGCACGCTGTGCGCAGATCTCCCGAGAACAGCTTTCGCGCAG 660
DB	
QY	661 GGGTTTCGATGCGCGCAACAATCATCATGATGAGTTGGGTTTCATCGACCCGCTGTGGCGCATC 720
DB	
QY	661 GGGTTTCGATGCGCGCAACAATCATCATGATGAGTTGGGTTTCATCGACCCGCTGTGGCGCATC 720
DB	
QY	721 AAGAGGTTCTTCCAGTCTGGGTTCAGAGGCCCTCTCTAGCGCAGAGCATCAAGCTCGTGGAC 780
DB	
QY	721 AAGAGGTTCTTCCAGTCTGGGTTCAGAGGCCCTCTCTAGCGCAGAGCATCAAGCTCGTGGAC 780
DB	
QY	781 GAGTTTCACCTACAGCGTGTATCCCGCGGAGGAAGCCGAGATCGTFCGAGGTTCCGGGCCAGC 840







```

QY 1584 GATGACCATCCTCTCCATGCGCGCACGGCCT 1613
Db 1485 GTGCTCAGCCTCTTATGAAGGGCGGCT 1514

RESULT 5
AA226874
ID AAX26874 standard; cDNA; 2181 BP.
AC AAX26874;
XX
XX
XX
XX 23-JUN-1999 (first entry)
DE Cytochrome P450 cDNA designated CYP86A5.
XX
XX Cytochrome P450; protein expression; yeast; ss.
XX
XX Triticum aestivum.
XX
XX Location/Qualifiers
FH Key 112..1737
FT CDS /*tag= a
FT
XX
XX FR2768748-Al.
XX
XX 26-MAR-1999.
XX
XX 24-SEP-1997; 97FR-00012094.
XX
XX 24-SEP-1997; 97FR-00012094.
XX
XX (RHON ) RHONE-POULENC AGROCHIMIE.
XX
XX Batard Y, Schalk M, Durst F, Werck RD;
XX
XX WPI; 1999-217499/19.
DR P-PSDB; AA101647.
XX
XX DNA for expression in yeasts - with codon changes based on yeast codon
XX usage.
XX
XX Claim 20; Page 21-23; 3lpp; French.
XX
XX The present sequence encodes a cytochrome P450, and was used to exemplify
CC the invention. The specification describes a DNA sequence that encodes a
CC protein of interest and contains regions with a high content of codons
CC poorly suited to yeasts, where a sufficient number of these codons are
CC replaced by corresponding codons suited to yeasts in these regions.
CC Yeasts transformed with vectors such DNA sequences can be cultured to
CC produce the protein of interest (especially a plant protein) or, when the
CC protein is an enzyme (such as cytochrome P450), can be cultured in the
CC presence of a substrate for the enzyme to produce a substrate conversion
CC product
XX
XX Sequence 2181 BP; 395 A; 722 C; 684 G; 380 T; 0 U; 0 Other;
XX
XX Query Match 14.2%; Score 271.6; DB 2; Length 2181;
XX Best Local Similarity 54.5%; Pred. No. 3.5e-42;
XX Matches 736; Conservative 0; Mismatches 524; Indels 90; Gaps 6;

QY 276 CACCTACATCGTGTGACCGGTTGAATGTCGAGCATCTCTCAGACTAACTTCAACCAATTA 335
Db 354 CACCCTACCTCGACCCCGCGCAACCTGGAGCACGTCCTTGAAGGCGCGCTTCGACACTTA 413
QY 336 CCCCAGGGAATCGTGACAGATCCTATGACGACGTGCTCTCGGTGAGCGGATCTTCAA 395
Db 414 CCCCAGGCGCCCTCTCTGCGACGGCTCTTCCGGGACCTGCTCGGAGCGGATCTTCAA 473
QY 396 CCGCGACGCGAGCTGTGTGAGGAAGCAGAGGAGCGGAGTTTCGAGTTCCGCTC--- 452
Db 474 TTCCGACGCGACACCTGCTCGCGCAGCGCAAGACGGCGCGGCTCGAGTTCAACCCCG 533
QY 453 CAAGAACCTGAGGGATTTTCAGCGCCATTTGTTTCAGAGAGTACTCCCTGAAGCTGTGCGG 512

534 CACGCTCCGGACGGCCATGTCCCGTGGGTCTCGCGCTCCATCCAGCGCGGCTCTCTGCC 593
513 TATACTGAGCCAGGATCCAAAGCAGGCAAAAGTTTGGACATGCGAGGAATTTATATAGAG 572
594 CATCTGGCCGACGCGGCCAAGGCGCAGGTGGATCTCCAGGACTCTCTCTCTCCG 653
573 GATGAGCTGGAATCAATCTGCAAGTTGGTGGGTTCGGGTTCGAGATCGGCACGCTGTCCG 632
654 CCTCACCTTCGACAAATCTCGGCGCTGGCCTTCGGCAAGGACCCGGAGACCTTCGCCA 713
633 AGATCTCCCGGAGAACAGCTTCGCGCAGGCTTCCGATGCGGCAACATCATCATCAGCT 692
714 GGGCTTCGCGGAGACGAGTTGCTCCGCTTCGACCGCGCCACCGAGGACGCTCAA 773
693 GCGGTTTCATCGACCGG-----CTGFGCGCATCAAGAGGTTCTTCCAGTGGGTTCAGA 746
774 CGGCTTCACTCTCCCGGAGTTCTGTGGCGCTGCAAAAAGTGGCTGGGCTTCGGCATGGA 833
747 GGGCTCTTAGCGCAGAGCATCAAGCTCGTGACGAGTTCACTACAGCTGATCCGCCG 806
834 GACCACGCTGACCAGCAGCATGGCCACGTCGACCACTGCTGCGCGCTCATCAAGAA 893
807 GAGGAAGGCGGAGATCGTCAGGTCGCGGCCAGCGCAACAGGAAAGATGAAGACGA 866
894 GCGCAAGCTCGAGTCGCGCGCGCAACGGCAATGGACACGCGCGGCGACGACGCA 953
867 CATCTGTACGCTTATCGAGCTGGGCGAGGCGCGGAGGAGGCGGCTTCGGGGA 926
954 CTGCTCTCCCGGTTTCATCGGAAGGTTCTACTCGGAG-----994
927 CGATAAGAGCTCGGGAGCTGTGTCAACTTCGTGATCGCGGCGGACACGACGCG 986
995 -----AGTCGCTCCAGACGTGGCTTCTTGGCTCGTGTCACCCACCTCGGTGGAGCGAAGAT 1109
987 GAGCAGCTGTCTGTGTTTACGCACATGGCCATGTCCCAAGCTGGTCTACCTCAAGCTGTGCTCAGAGCT 1046
1050 CTGCGGCTCTCTGTTCTTCTGCTGCTGTCGTCACCCACCTCGGTGGAGCGAAGAT 1109
1047 GCGCGCGAGCTGTGCGGCTTCGAGGCGGAGCGCGGCGGAGGAGGCGGTTCAGCTCGT 1106
1110 CGTGGCGAGCTCTGCTC-----CGT 1130
1107 GCTCTGGCGGCGGCTGACGCGCGACGAGGCTTCGCGCGCGGCTGGCGAGTTTCG 1166
1131 TCTCGCGCGTACGCGGCGCCCATGACCCGCAAT-----GTGGCTGGC 1175
1167 GGGCTCTCTCACTTACACAGCTCTCGCAAGCTGGTCTACCTCAAGCTGTGCTCAAG 1226
1176 GGAGCCCTTCACTTTCGAGGAGCTCGACCGCTGTCTACCTCAAGGCGGCTGTGCGA 1235
1227 GAGCTCGGCTGTACCGCGCTCTCTGAGGACCCCAAGGAGTCTCTGGAGGACGAGCT 1286
1236 GACCTCGGCTTACCCCTCGCTCCCGGAGGACTCAAGACGCTGCTCGGAGGACTA 1295
1287 GCTCGCGAGCGGACGAAAGTGTAGGCGCGCGGATGCTGACGCTACGCTCTACTCGAT 1346
1296 CTTCCCGAGCGACCTTCTGCGCGCGGCTGTGCTGCTACCTTCTTCAATACTCGG 1355
1347 GGGCGGATGAGTACAACTGGGCGCGCGAGCGGCGGAGCTTCCGCGCGGAGCGGTGAT 1406
1356 GGGCGGATGAAGGGGTGTGGGGGAGGAGCTGCTCGAGTTCCGCGCGGAGCGATGGCT 1415
1407 CAAAGAGTGGCG-----GTTCCGCAAGCTGCTGAGTTCAGGTTCCAGGCTTCCAGG 1463
1416 GTCGCGCGACGCGCACCAAGTTCGAGACGACGACTCGTACAAAGTTCTGGGCTTCAAGC 1475
1464 GGGCGGAGGATCTGCTGGCAAGGACTCGGCTGCTGAGATGAAGTGGCGCTGGC 1523
1476 CCGCGCGAGGCTGTCTTGGCAAGGACTAGCTTCTGAGATGAAGAACATCGCGG 1535
1524 CATCTCTTCCGCTTCTACAGCTTTCGCGCTGCTGGAGGCGCACCCGCTGCGATCCGAT 1583

```

Db 1536 GAGCGTGTGCTTCGGCCACCGCTGACCGTGGCGCGGCCACCAGGTGAGCAGAGAT 1595  
QY 1584 GATGACCACTCTCTCATGGCGACGGCCT 1613  
Db 1596 GTCGCTCAGCTCTTCATGAAGGCGGGCT 1625  
RESULT 6  
ID ADA70596 standard; DNA; 1527 BP.  
XX ADA70596;  
XX 20-NOV-2003 (first entry)  
XX Rice gene, SEQ ID 3919.  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.  
XX Oryza sativa.  
XX OS  
XX WO2003000898-A1.  
XX 03-JAN-2003.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPT; 2003-175290/17.  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX Claim 6; SEQ ID NO 3919; 999pp; English.  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
XX illustrate the invention.  
XX Sequence 1527 BP; 280 A; 473 C; 494 G; 280 T; 0 U; 0 Other;  
S  
Query Match 11.9%; Score 226.4; DB 7; Length 1527;  
Best Local Similarity 54.2%; Pred. No. 1.3e-33;  
Matches 678; Conservative 0; Mismatches 476; Indels 96; Gaps 7;  
QY 266 TCATCTTCCATACCTACATCGCTGAGCCCGTGAATGTCGAGCATGCTCTCAGAGCTAACT 325  
Db 239 TCACCGCGCGCTCATACCGCCAAACCCGCGCAATGTCGAGTATACGCTGAAGACTTAAT 298  
QY 326 TCACCAATTAACCCAGGGAATCGTGTACAGATCTCTACATGACGCTCTCTCGGTGAGC 385  
Db 299 TTGGTAATACCCCAAGGCGAGCTCGCGTGTCCATGTCTGTCTGCTGCTCTCTCGGCCATG 358  
QY 386 GCATCTTCAACCGCGAGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTTTCAGT 445  
Db 359 GCATCTTCAACTCCGACGGGAGCACTGGCAGTGGCAGCGGAAGCGGCCAGCTACGAGT 418

QY 446 TCGCCTCAGAGAACTGAGGGATTTC---AGCGCAATTGTGTTTCAGAGAGATACCTCCCTGA 502  
Db 419 TCACAAGCGCTCGCTGAGGAACTTCGTGGTGACACCGTCAGGTCCGAGGTCTGTCGAGA 478  
QY 503 AGCTGTGGGTATATGAGCCAGGCAATCAGGCGAGGCAAGTTTGTGACATGCGAGAAC 562  
Db 479 GGCTGTGCGCTGTGAGCGCGGAGCGAGCGCGCGGACGCTGACGCTGTCAGGACG 538  
QY 563 TTTATACAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTTCGGGGTTCAGAGATCGGCA 622  
Db 539 TCGTGGAGCGCTTCGGGTTTCGACAACTGTCATGTGCTTCGCTTCGACAGGA----- 590  
QY 623 CGCTGTGCGCCAGATCTCCCGGAGAAAGCTTCGCGAGGCGCTTCGATGCGCGCAATCA 682  
Db 591 ----CCCGCGTGCCTCGCCGAGGACAGCATGCG---CTCGCCCGCAGAGCGGGAATTCA 643  
QY 683 TCATACGCTGCGGTTTCATCGACCGCTGTGCGGATCAAGAGGTTCTTCCAGCTCGGT 742  
Db 644 TCGCGCGCTTCAGTTCAGCGCGCAATGCGGTATGAGACCGGTTTCATGTGCGCGTCAAGT 703  
QY 743 CAGAGCCCTCTAGCGCAGAGCATCAAGCTCGTGACGAGTTCACCTACAGCGTTCATCC 802  
Db 704 CGCGGTGGCGCTTCAGAGGCTGTTTCAATGAGCCCGAGGCGAGATCGGAGGCGCG 763  
QY 803 GCGGAGGAGGCGCGAGATCGTTCAGGTCGCGGCGAGCGGCAAGAGAGAGATGAGC 862  
Db 764 TCGCCAGATCCACGCTTCGCGGAG---CGGATCGTCCGCGAGCGCAGGAGAGGGG 820  
QY 863 ACAGATCTCTGTCACGCTTCATCGAGCTGGCGAGCGCGCGGCGAGCGCGCGGCTTCG 922  
Db 821 AGGCGCGCTGGCGCGCAGTTCGCTGCGGCTTCGCGCGGAGCGCGGCGGATCACA 880  
QY 923 GGCAGCATAGAGCTCCGCGAGCTGCTCAACTTCGTGATCGCGCGCGGAGCAGCA 982  
Db 881 GCGAGC---AGAGCTCCGTCAGCTGTGTGACAACTTCCTCATCGCGCGCGAGCAGCA 937  
QY 983 CGGCGAGCAGCTGTGCTGTTTCAAGCATGCGCATGTCCTCCACCGCGAGCTGCGCGAGA 1042  
Db 938 GCTCGAGCGGCTTGAACCTGTTCTTCTGCTACTGTCCGCGCGCGCGCGAGCTGAGGACA 997  
QY 1043 AGTTCGCGCGGAGCTGTGCGGTTTCAGGCGGAGCGCGCGCGGAGGCGGCTCAGCG 1102  
Db 998 AGATCGTACGCGAGATCCAGCA-----1020  
QY 1103 TCGTGTCTCGCGCGCGCTGAGCGCGACGACAGGCGTTTCGCGCGCGCGGTGCGCAGT 1162  
Db 1021 --GTGCGCGCGCTTCGCTGGGCAAGGCGTTCGCTGCTTGTGCTTGTGCTTGTGCTGCT 1057  
QY 1163 TCGCGGCGCTCTCTCACTACAGAGCTTCGCGAAGTGGTCTACTCAGCGCTGCGTCA 1222  
Db 1058 -----TCAACTTGGACGAGCTGCTGACATGCAATACCTCCATGCGCGCATCA 1105  
QY 1223 CCGAGACGCTCGCGCTGTACCCCGCTTCCTCAGGACCCCAAGGGATCTCTGAGGACG 1282  
Db 1106 CCGAGTCCATGCGCTGTACCCACCCGTTGGCCATGGAATCTGCAAGCTGCAAGGAGGAGC 1165  
QY 1283 ACGTGTCTCGCGACGCGAAGGTTGAGGCGCGCGGCGGATGCTGACGTACGTGCGCTTACT 1342  
Db 1166 ACTTCTCTCGGACGCGACGCTTCGTGGGGAAGGCTGCTTGTGCTTGTGCTTGTGCTTGTG 1225  
QY 1343 CGATGGGCGGATGAGATCAACTTGGGCGCGCGAGCGCGAGCTTCGCGCGGAGCGGT 1402  
Db 1256 CGATGGCAGCGGTGGAGGACATCTGGGCGCGGAGCTGTGAGGAGTTTCAAGGCGAGCGGT 1285  
QY 1403 GGATCAAGAGGATGGCGGCTTCGCAACGCTCGCGCTTCAAGTTTCAAGTTTCAAGTTTCCAGG 1462  
Db 1286 GGTAGACGAGCGCGCGCTTCGCGCGGAGAGCGCGCTTCAAGTTTCAAGTTTCCAGG 1345  
QY 1463 CGGCGCGGAGGATGCTGCTGCGCAAGGACTCGCGCTTACTGTCAGATGAAG 1512  
Db 1346 CGGCGCGGAGGATGCTGCTGCGCAAGGAGTGGCTTACATACAGATGAAG 1395



RESULT 7	QY	475	GCATTTGTTTCAGAGAGTACTCCCTGA---AGCTGTGGGTATACTGAGCCAGGCATCC	531
ADA70663	DB	469	TACAAGATCATGCCCCAGAGGTGAGCCCGCTCATGCCGTGCTCGCCGACGCCGCC	528
AC ADA70663;	QY	532	AAGCAGGCAAGTTGTGGACATCAGGAACTTTACATGAGGATGACGTGGATCCATC	591
DT 20-NOV-2003 (first entry)	DB	529	GACCGGGCGCGTCTCGATCTCGAGACGTGTTCCGCCGCTTCGCCCTTGACAATC	588
DE Rice gene, SEQ ID 3986.	QY	592	TGCAAGTTGGTTCCGGGTCGAGATCGGACAGTGTGCGCAGATCTCCCGGAGAACG	651
KW Plant; bacterial infection; fungal infection; viral infection; rice;	DB	589	TGCAAGATCTCTTCGGCTCGACCCCGCTCGACCGGAAATGCCCCGTGTCCGAG	648
KW gene; ds.	QY	652	TTCCGCGAGGGTTTCGATGCGCCCAACATCATCATACGCTCCGCTTCATGACCC	707
OS Oryza sativa.	DB	649	CTGCCGACGCTTCGACGCGCGCTCGCGCTCTCCGCCATGCGTGGGAGCGCGTTG	708
PN WO2003000898-A1.	QY	708	-----GCTGTGGCGCATCAAGAGTTTTCACGTCGCGGTGAGAGCCCTCTAGCAG	762
PD 03-JAN-2003.	DB	709	CGTTGCTGTGAGGGGAAAGCGGTTCTCAACGTCGCGTGGAGAGGAGCTCAGGAG	768
PF 22-JUN-2001; 2001WO-IB001105.	QY	763	AGCATCAAGCTCGTGAGAGGTTCACTACAGCGTGATCCCGCGGAGAGCCGAGATC	822
PR 22-JUN-2001; 2001WO-IB001105.	DB	769	GCCATCAAGGTCGTCGACGAGCTCGCGCGGATGATCCCGGAGCGCGAGCTGGGC	828
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.	QY	823	GTCGAGGTCGGGCGCAGCGGCAACAGGAGAGATGAAGACACGATCTGTCAACGTT	882
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;	DB	829	GTCC-----GCAGCAGCCACGACCTCTCTGCGCGTTC	861
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;	QY	883	ATCGAGCTGGGCGAGCGCGGCGACGAGCGCGCGCTTCGGGAGAGTAAGAGCTCCG	942
PI WPI; 2003-175290/17.	DB	862	AT-----GGCTCCACCGCGCTCGACGACGCGCGCGGAGACGACAACTTCCTCCG	912
XX Identifying at least one gene involved in plant resistance or response to	QY	943	GACGTGCTCTCAACTTCTGATCCCGCGGCGGACAGGCGGACGAGCTGTCTGCTGG	1002
XX pathogenic infection for conferring resistance or tolerance to a plant to	DB	913	GACATGCTGCTCAGCTTCTCTCGCGCGGCGGACACGCTGTCCACCGGCT-----	965
XX bacterial, fungal or viral infection by determining or detecting plant	QY	1003	TTACGACACATGGCCATGTCCACCGGACGTGGCGGAGAAAGCTGGCGCGAGCTGTGC	1062
XX gene expression.	DB	966	-----CACCACGCTGTTC	978
XX Claim 6; SEQ ID NO 3986; 899pp; English.	QY	1063	GGTTTCGAGCGGAGCGCGCGGAGGAGGCGTCACTGCTGTGCTGTGCGGCGGCT	1122
XX The present invention relates to a method (M1) for identifying genes	DB	979	ATGCTCTCTCAAGAAACCCGAGTCCGCGCGCCATCGCGCGGAGCGGAGCGCGC	1038
XX involved in plant resistance or response to pathogenic infection. M1	QY	1123	GACCGCGACACAGGCGTTCCCGCGCGGTCAGTTCGCGGCGCTCTCTCACTAC	1182
XX comprises identifying a gene whose expression is significantly altered in	DB	1039	GACGCGCGGAGACGCGCGCGC-----CATCAGCTAC	1071
XX the incompatible interaction of plant gene expression relative to	QY	1183	GACAGCTCGGCAAGCTGTCTACCTCCACGCTGCTCACGAGACGCTCCGCTGTAC	1242
XX expression of the gene in an uninfected plant, in a mutant plant that	DB	1072	GAGCCTCAAGGCGCTCCACTACACCGCGCTGCTGCAGAGAAACATGCGGCTGTTC	1131
XX does not express a gene associated with response to pathogenic infection,	QY	1243	CCCGCGCTCTCAGACGCCCAAGGGATCTCGAGAGACGAGTGTGCGGAGCGGACG	1302
XX or in a corresponding incompatible or compatible interaction. (M1) is	DB	1132	CCCGCGGTGAGTTGACTTCAAGTTTTCGCGCGCGCGACGCTGTCTCCCGACGAC	1191
XX useful for conferring resistance to resistance or tolerance to a plant to	QY	1303	AGGTGAGGCGCGCGGATGAGTACGTGCTGCTCTCTACGATGGGCGGATGAGTAC	1362
XX bacterial, fungal or viral infection. The present sequence was used to	DB	1192	TACGTGCGCGGCGACGCGCGCTCATGTACACCCCTACGCCATGGGCGGATGCCGAC	1251
XX illustrate the invention.	QY	1363	AACGCGCGCGCGCGGAGCTTCCGCGCGGAGCGGTGGATCAACAGAGATGGCGCG	1422
XX Query Match 11.2%; Score 214.4; DB 7; Length 1563;	DB	1252	ATCTGGGCGCGCGCTTCCGCGCGGCTTCCGCGCGGCTGCTCACCGCGCGCGCG	1311
XX Best Local Similarity 52.5%; Pred. No. 2.4e-31;	QY	1423	TTCCGCAACCGTGC---CCGTTCAAGTTTCAAGGCTTCCAGCGCGGCGCGAGTCTGC	1479
XX Matches 738; Conservative 0; Mismatches 536; Indels 132; Gaps 7;	DB	1312	TCGTTCTCCCGCGAACCCTGATAAATACCCGTTGTTCCAGGCGCGCGCTTC	1371
QY 175 ACGTGTGAGAGCTGAGGACTACACCGGATGACGAGTGGCTGTGCGGTACTCTCA	QY	1480	CTGGGCGAGGACTCGGCTGCTGCGATGAAGATGCGCTGCGCTCTCTTCGCGTTC	1539
DB 169 ACGGGTCTGGGCGAGGAGTTACCAACCTCGGCGAGTGTACGCGACCTCTCGC	DB	1372	CTCGGCAAGGAGTCTCGCTGTCACCGAGATGAAGCGGCGGAGCTGCGCGCTG	1431
QY 235 CGGCACAGGACAGTGACCGTTCGACATGCGTTCACTTCTACACCTACATCGTACCCG	QY			
DB 229 CGTTCGCGAGCGGAAACGTTCCAGTCCAGCTCTCGGCTGACCGTCAACGCGAAC	DB			
QY 295 GTGAATGTGAGCATGCTCTCAAGACTAACTTACCAATACCCCAAGGGAATCGTGTAC	QY			
DB 289 GCGAAGTTCGAGCATGCTCAGGACGCGGTTCCAGCACTTCCCAAGGGAGGCCCTTC	DB			
QY 355 AGATCTTACATGAGCTGCTCTCGTACCGCATCTTCAACGCGAGCGGAGCTGTGG	QY			
DB 349 GCGGCGCTCCTCGGCGACTCTCTCGGCGACGCGATCTTCAACGTCGAGCGGCGAC	DB			
QY 415 AGGAGGAGAGGAGACGCGGAGTTTCGAGTTCGCTCCAGAACTTCAGGAGTTTCAGC	QY			
DB 409 GCGCCACGCGGAAGATGGCCAGCTCGAGCTCGGAGGCTCGGCTCGCTTCACGCC	DB			



[illegible]

RESULT 9	
ABZ14176	
ID	ABZ14176 standard; DNA; 1473 BP.
XX	
AC	ABZ14176;
XX	
DT	21-JAN-2003 (first entry)
XX	
DE	Arabidopsis thaliana stress regulated gene SEQ ID NO 1981.
XX	
KW	Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200216655-A2.
XX	
PD	28-FEB-2002.
XX	
PF	24-AUG-2001; 2001WO-US026685.
XX	
PR	24-AUG-2000; 2000US-0227866P.
XX	
PR	26-JAN-2001; 2001US-0264647P.
XX	
PR	22-JUN-2001; 2001US-0300111P.
XX	
PA	(SCRI ) SCRIPPS RES INST.
XX	
PI	(SYGN ) SYNGENTA PARTICIPATIONS AG.
XX	
PI	Harper JF, Kreps J, Wang X, Zhu T;
XX	
XX	WPI; 2002-304127/34.
DR	
XX	
PT	Identifying a stress condition to which a plant cell has been exposed and
XX	
PT	producing plants with increased tolerance to these abiotic stresses.
XX	
PS	Claim 144; SEQ ID NO 1981; 577pp + Sequence Listing; English.
XX	
CC	The invention relates to identifying a stress condition to which a plant
XX	
CC	cell has been exposed, comprising: (a) contacting nucleic acid
XX	
CC	representative of expressed polynucleotides in the plant cell with an
XX	
CC	array or probes representative of the plant cell genome; and (b)
XX	
CC	detecting a profile of expressed polynucleotides in the plant cell
XX	
CC	characteristic of a stress response. The method is useful in the
XX	
CC	production of transgenic plants, cells and seeds and in producing plants
XX	
CC	with increased tolerance to abiotic stress. The present sequence is that
XX	
CC	of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
XX	
CC	in methods of the invention. Note: The sequence data for this patent is
XX	
CC	not represented in the printed specification but is based on sequence
XX	
CC	information supplied to Derwent by the European Patent Office
XX	
SQ	Sequence 1473 BP; 443 A; 285 C; 343 G; 402 T; 0 U; 0 Other;
Query Match	8.7%; Score 165.4; DB 6; Length 1473;
Best Local Similarity	49.5%; Fred. No. 4.8e-22;
Matches 642; Conservative	0; Mismatches 571; Indels 84; Gaps 5;
QY	273 CTACACCTACATCGCTGACCCGGTGAATGTGGAGCATGTCTCAAGACATACTTCAACAA 332
DB	174 CGAGTATTCATCGAGATCTCTCGAAGCTGGAGCATATTTCTCAAGACAAGATCCATAA 233
QY	333 TTACCCCAAGGGAATCGTGTACAGATCTTACATGGAAGTGTCTCTCGGTGACGGCATCTT 392
DB	234 CTATAGTAAAGGACCCGTTGGTACAGTGAATCTTCGGATCTCTTCGGACATGGGATCTT 293
QY	393 CAACCCGACGCGAGCTGTGAGGAGCAGAGGAGAGCGGAGTTCGAGTTCGGCTC 452
DB	294 CGCTGTGTATGAGAGAAATGGAAACAAAGAGGAGCTCTGTGAGCTTTGAGTTCTCCAC 353
QY	453 CAAGAACCTGAGGGATTTTCAGCGCAATTTGTGTTTCAGAGAGTACTCCCTGAAGCTGTCCGG 512
DB	354 TAGAGTTTAAAGAAATTTTAGCTACTCTCTCTTTTCGACAAAGTCTTCTAAGCTTTGTCGG 413

QY	513 TATACTGAGCCAGGCATCCAAAGCCAGCAGCAAGTTGTGGACATCGAGAACTTTACATGAG 572
DB	414 TTTTATGCGGAGTTTGTCTCTCTCTCTGAAATCTTTTGAATTTTTCAGATATGTTGATGA 473
QY	573 GATGACGCTGGACTCCATCTGCAAGGTTGGGTTGCGGTTGAGATCGGACGCTGTGCGCC 632
DB	474 ATGTACATTTGGAATCTTTAAAGTTGGAATTCGGTGTGGAGTTAGGATGTTCTAGATGG 533
QY	633 AGATCTCCCGA--GAACAGCTTCGCGCAGGCGTTTCGATGCGCGCCCAACATCATCATCAC 689
DB	534 GTTTAGCAAGAAGGGAAGAGTTTCATGAAGGCTTTTGTATGAAGGCAACCGCGCAACTAG 593
QY	690 GCTCGGTTTCATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCACGTCGGGTTCAGAGGC 749
DB	594 TTCAGGGTCAACCGCCGTTTTCGAAAGCTGAATGTTTCTTAACATTTGGATCAGAGTC 653
QY	750 CCTCCTAGCGCAGAGCATCAAGCTCGTGAGAGTTTCACTTACAGCGTGTATCCCGCGAG 809
DB	654 AAGACTCAAGAAGAGCATTTGCTATTATAGACAAGTTTGTCTATAGTCTCATTTACACTAA 713
QY	810 GAAGCCGAGATCGTCGAGGTCGCGGCCAGCGCAACAGGAGAGATGAAGACGACAT 869
DB	714 AAGGAAAGAGCTTTCCAAAGGAACAGACA-----CTTCTGTTAGAGAGCAT 761
QY	870 CTTGTACCGGTTTCATCGAGCTGGCGAGCCGCGGAGCAAGCGCGCGCTTCGGGAGCA 929
DB	762 CTTATCGAAATTTCTTCTCGAGAGTGAGAAAGATCCGGAGA-----ACATGAATGA 812
QY	930 TAAAGAGCTTCGCGACGTTGGTCTCAACTTCGTGATCGCGCGGCGGAGACAGCGCGCAC 989
DB	813 TAAGTACCTGAGGAGATATCAITTTGAATGTTATGTTGTTGTTGTAAGGACACAAACCGCTGC 872
QY	990 GACGCTGTGCTGTTTCAAGCATGCGCATGTCCTCCACCGCGAGCTGGCGCGAAGACTGCG 1049
DB	873 ATCACTCTCTTGGTCTTGTATCATGCTCTGCAAAACCCACTTTGTTCAAGGAA----- 926
QY	1050 CCGCGAGCTGTGCGCTTCGAGGCGAGCGCGCGCGAGGAGGCGTCACTGCTGCT 1109
DB	927 -----AATCGTTCA 935
QY	1110 CTGCGCGCGCTGAGCCGACGACAGCAGTTCGCGCGCGCTGGCGAGTTCGCGG 1169
DB	936 AGAAATCAGAGATGTGACATCAAGTCACGAGAAACCAACCGATGTAAATGGTTTCATTGA 995
QY	1170 CTTCTCCTCCTACGACAGCTTCGGCAAGCTGTCTACCTCCAGCGCTGGTCAACGAGAC 1229
DB	996 AAGTGTAAACGAGAGAGCTTTGCTCAGATGCAATCTCCATGCGGCTTTGTCTGAGAC 1055
QY	1230 GCTCCGCTGTACCCGCGCTCCCTCAGGACCCCAAGGGGATCCTGGAGAGCAGCTGCT 1289
DB	1056 TATGAGGCTTTTACCCACCTGTGCTGAGCACAATGAGGTGTGCAAGAAATGATGACGTTCT 1115
QY	1290 GCGGAGCGGACGAAGTGTAGGCGCGCGGATGTGACGTAGCTGCTTCTGATGGG 1349
DB	1116 TCCAGATGGAATAGAGTGTAGCAAGAGGATATATCTACTATCATATCTCTATGCAATGGG 1175
QY	1350 GCGGATGAGCTACAACTGGGCGCCCGACGCGGCGAGCTTCGCGCGCGAGCGGTGTGATCAA 1409
DB	1176 TAGATGACTTACATTTTGGGCTCAAGATGCTGAGGAATTCAGGCAAGAGATGGCTCA- 1234
QY	1410 CGAGGATGCGCGCTTCGCAACCGGTCGCGTTCAAGTTTCAAGGCTTCAGGCGGTCCAGGCGGGGCC 1469
DB	1235 --AGGACGGCGTATTCACACCCGAATCAAAATTTCAATTTTCAAGCTTTTATGCTGTGTC 1292
QY	1470 GAGGATCTGCTGGCAAGGACTCGGCTACCTTCAGATGAAGATGGCGTGGCCATCTCT 1529
DB	1293 AAGAAATCTGTATTGGCAAGGATTTCCGCAATCCGCGCAGATGAAGATAGTATCAATGGCACT 1352
QY	1530 CTTCCGCTTCTTACAGCTTTCCGGCTGCTGGAGGGGCAAC 1566
DB	1353 TCTTCACTTTCTTTCGCTTCAAAATGGCTGTGATGAGAAC 1389

RESULT 10  
ADA70631  
ID ADA70631 standard; DNA; 1506 BP.  
XX  
ADA70631;  
AC  
20-NOV-2003 (first entry)  
XX  
DE Rice gene, SEQ ID 3954.  
XX  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.  
XX  
XX Oryza sativa.  
OS  
XX  
XX WO2003000898-A1.  
FN  
XX  
XX 03-JAN-2003.  
PD  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
PF  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
PR  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
PA  
XX  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
PI  
XX  
XX WPI; 2003-175290/17.  
DR  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
XX Claim 6; SEQ ID NO 3954; 899pp; English.  
PS  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
XX Sequence 1506 BP; 208 A; 540 C; 521 G; 237 T; 0 U; 0 Other;  
SQ  
Query Match 8.3%; Score 159; DB 7; Length 1506;  
Best Local Similarity 51.0%; Pred. No. 7.9e-21;  
Matches 716; Conservative 0; Mismatches 530; Indels 159; Gaps 8;  
QY 176 CGGTGGAGAGTGGAGAACTACCAACCGGATCCAGACTGGTGTGGGTACCTGTCTAC 235  
Db 146 CCGGGTCGTGGAGAGGAGGAGTTCCGCAACCTCGGCGACTGGTACCGCACTGTCCGCC 205  
QY 236 GSCACAGGACAGTACCGTCCGATCGGTTTCACTTCCTACACCTACATCGCTGACCCGG 295  
Db 206 GCTCGCGACGCGGACCGTCCACGTCACGTCCTCGGTGACCGTCCGCTACCGCGAACC 265  
QY 296 TGAATGTCAGACATGTCTCAAGACTTACCTTCAACCAATTACCCCAAGGGAATCGTGTACA 355  
Db 266 CGAAGCTGAGTACATGCTCAAGACGCGCTTCGACAACTTCCCAAGGAGGCGGTTCCG 325  
QY 356 GATCTTACATGAGCGTCTCTCGGTGACGCGCATTTCAACGCGGACGCGAGCTGTGGA 415  
Db 326 CCGCGTCTCTCGCGACCTCTCTCGGCGCGGATCTTCAACGTCGACGCGGACGCGTGGC 385  
QY 416 GGAGCAGAGGAGAGCGGAGTTTCGAGTTCGCTTCCAGAACCTGAGGGATTTC---A 472  
Db 386 GCACACGAGAGATGGCGAGCTCGAGCTGGGGAGCGTCCGCGTGAGATCTTACGCGT 445

QY 473 GGGCCATTGTTTTCAGAGAGTACTCCCTGAAGCTGTCGGGTATCTAGCCAGGCAATCCA 532  
Db 446 ACAAGATCGTCGCCCCAGGAGGTGGAGCCCGCTCATGCCGGTGTCTGCCAACGCCGCCG 505  
QY 533 AGGCAGGCAAGTGTGTGACATGCAAGAACTTTTACATGAGGATGACGCTGGATCTCATCT 592  
Db 506 ACAGCGGCGCGTGTGTGACCTGCGAGGAGTGTTCGGCGCTTCGCTTCGACACCATCT 565  
QY 593 GCAAGTGTGGGTTCGGGTCGAGATCGGCACGCTGTGCCAGATCTCCCGGAGACAGCT 652  
Db 566 GCAAGATCTCTTCGGCTTCGACCCGCGCTGCTGACCCGGAGATGCCGCTGTGGAGC 625  
QY 653 TCGCGAGGCGTTCGATGCCGCCAACATCATCATCGCTCGGTTTCATCGACCC--- 707  
Db 626 TCGCGAGCGTTCGACCGCGCTGCGGCTGTCCGCTATGCTGCGCGCGCGGCTCGC 685  
QY 708 ---GCTGTGGCGCATCAAGAGTTCCTCCAGTGGGTTCAGAGGCCCTCTTACGCCAGA 763  
Db 686 CGTGTGTGGAAGATGAAGCGTTTCTCAACGTGGGTTCGAGAGGAGCTCAAGAAG 745  
QY 764 GCATCAAGTGTGTGAGAGTTCACCTACAGGTGATTCGCCCGGAGAGGCCGAGATCG 823  
Db 746 CCATCAAGTGTTCGACCGCGCTCGCGGCGCGATGATTCGGGAGCGCGGAACTTGGCG 805  
QY 824 TCGAGTTCGGGCGCGAGGAGATGAAGACAGGAGATGAAGACAGCATCTCTGTACGGTTCA 883  
Db 806 TCG---CGACAGCCACGACCTCTCTGTCGGGTTC 838  
QY 884 TCGAGTTCGGGCGCGCGAGCGCGCGCTTCGGGGACGATAGAGCCTTCGGG 943  
Db 839 ---TGGCTCTCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 878  
QY 944 ACGTGTGCTCAACTTCGTGATCGCGGGGCGGACAGCGCGAGCGAGCTGTCTGTGT 1003  
Db 879 ---CAACTTCTCTCTCGCGCGCGGACACGCTGCTCTCGCGCT--- 920  
QY 1004 TCACGCACATGGCCATGTCCACCCCGAGCTGGCGGAGAGCTGGCGCGGAGCTGTGCG 1063  
Db 921 ---CACCACCTCTGTTC 934  
QY 1064 CGTTCGAGGCGGAGCGCGCGGAGGAGGCGTCAAGCTGTGTCTGTGGCGGCGCGCTG 1123  
Db 935 TGATCTGTCCAGAACCCCGAGCGCGCGCATCGCGCGGAGCGCGCGCGCGCGCGCG 994  
QY 1124 AGCGCGAGCAAGGCGTTCGCCCGCGGTGGGCGAGTTTCGGCGGCTCTCTCACTACG 1183  
Db 995 CCGCGGAGAGC-----GCCGCGGTTCAGTACG 1021  
QY 1184 ACAGCTTCGGCAAGCTGTCTACCTCCAGCGCTGCGTCACCGAGACGCTCCGCTGTACC 1243  
Db 1022 AGCACCTGAGCGGCTGAATACACCCAGCGCGTGTGTAGAGAACATGCGGTGTTC 1081  
QY 1244 CCGCGCTCTCTCAGACCCCAAGGGGATCTCGAGGAGCGAGTGTCTGCGCGAGCGGAGCA 1303  
Db 1082 CGCGGTGCGAGTTCAGTTCCTCAAGTTCTCGCGCGCGCGCGAGCTCTCCCGACGCGCACT 1141  
QY 1304 AGGTGAGGCGCGCGGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1363  
Db 1142 ACGTGACGCGGCGCGCGCGCTCATGTACCAACCCCTACGCCATGAGCGCGATGCCGCA 1201  
QY 1364 ACTGGGCGCGCGAGCTTCGGCGCGGAGCGTTCGGCGCGGAGCGGTGG---ATCAACGAGGATGCG 1420  
Db 1202 TCTGGGCGCGCGATGCGACCGCTTCGGCGCGGAGCGGTGGTAAACCGCGCGCGCGCG 1261  
QY 1421 CGTTCGCGCAACGCTGCGCTTCAAGTTTCAAGGTTTCAAGGTTTCAAGGTTTCAAGGTTTCA 1480  
Db 1262 CGTTCGTCGCGGAGAGCTCTTCAAGTACCGCGTTCAGGCGCGCGCTTCGCGGTGTC 1321  
QY 1481 TGGGCAAGAGCTTCGGGCTTCTGCGAGATGAAGATGGCGCTGGCCATCTCTTCCGCTTCT 1540  
Db 1322 TCGGCAAGAGCTTCGCCATCACCGAGATGAAGGCGGTTCAGCGTTCGCCCTGTGAGGGCAT 1381

QY	1541	ACAGCTTCGGGCTGCTGGAGGGCA	1565	QY	416	GGAAAGCAGAGAACGCGGAGTTTCGAGTTCCGCTCCAGAACTGAGGAGTTTCAGCG	475
Db	1382	TCGACGTCGAGGTCGTCGGCGAGAA	1406	Db	386	CCACACGCGGAGATGCGCAGCCTCGAGCTGGGAGCGGTGCGGTGAGATCCTACGCGT	445
RESULT 11				QY	476	CCATTGTGTTTCAGAGAGTACTCCCTGAA--GCTGTGCGGGTATATCTGAGCCAGGACATCCA	532
ID	ADA48303	standard; DNA; 1395 BP.		Db	446	ACAAGATCGTCGCCAGGAGGTGGAGGCCCGCTCATGCCGCTGCTCGCCAAACGCGCGG	505
AC	ADA48303;			QY	533	AGGAGGCAAAAGTTGTGACATAGAGAACTTTACATGAGGATGACGCTGACCTCCATCT	592
XX	20-NOV-2003	(first entry)		Db	506	ACAGCGCGCGCTGTGTCACCTGACCTGACGAGTGTTCGCGCTTTCGCTTTCGACCACTCT	565
DE	Rice gene conferring disease resistance in plants.			QY	593	GCAGGTTGGTTCGGGTTCGAGATCGGCACGCTGTGCGCAGATCTCCCGAGAACACGCT	652
XX	disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.			Db	566	GCAGATCTCTCTTCGGCTTCGACCCGCGCTGCTCGACCGGAGATGCCGCTGTGCGAGC	625
XX	Oryza sativa.			QY	653	TCGCGCAGGCGTTTCGATGCGCCAAACATCATCATCATCATCATCATCATCATCATCAT	707
PN	WO2003000906-A2.			Db	626	TCGCGCAGCGTTTCGACCGCGCTGCTGCGCTGCTGCGCTGCTGCGCTGCTGCGCTGCG	685
XX	03-JAN-2003.			QY	708	-----GCTGTGCGCATCAAGAGTTCTTCACGCTCGGCTCAGAGGCCCTCTCTAGCGCAGA	763
XX	21-JUN-2002; 2002WO-1B002453.			Db	686	CCTTCTGTGCGAAGATGAAGCGGTTTCTCAACGCTCGGTCGAGAGGAGCTCAAGAAGG	745
XX	22-JUN-2001; 2001US-0300112P.			QY	764	GCATCAAGCTCGTGACGAGTTTCACTACAGGCTGATCCGCGGAGGAGGAGGAGGAGGAG	823
PR	26-SEP-2001; 2001US-0352277P.			Db	746	CCATCAAGCTCATGACGCGGCTCGCGCGCATGATCC-----GGAGGCGCGGAGC	799
PR	22-MAR-2002; 2002US-0366535P.			QY	824	TCGAGTTCGGGCGCGCAACAGAGGAGAGATGAAGACGACATCTCTGTTCACGTTTCA	883
XX	(SYGN ) SYNGENTA PARTICIPATIONS AG.			Db	800	TTGGGCTCGGAACAGCCACGACCTCTCTGCGGTTTCATGCGCATCTGCTGTAAGCTTCC	859
XX	Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;			QY	884	TCGAGTTCGGGCGCGCGAGCGCGGCGGCTTCGGGAGCAGTAAAGAGCTCCGCGG	943
PI	Katagiri F, Xreps J, Provart N, Ricke D, Zhu T;			Db	860	TCCTCGCGCGCGGACACGCTGCTCTCGCGCTACCACTCTGTTTCATGA---TCTCT	916
XX	WPI; 2003-184052/18.			QY	944	AGTGTGTCTAACTTCGTGATCGCGCGGCGGAGACACGACGCGGAGCAGCTGTCTGTGT	1003
DR	P-PSDB; ADA48304.			Db	917	CMAGAACCCGACGTTGGCGCGCGCATGCGCGGAGGCGCGCGCGCGCGCGGAG	976
XX	New polynucleotide comprising a plant nucleotide sequence having an open			QY	1004	TCACGCACATGGCCATGTCCACCCCGGAGCTGCGCGGAGGAGCTGCGCGGAGCTGCGG	1063
XX	reading frame that encodes a polypeptide associated with disease			Db	977	GCSCCGCGCTCAGCTACGAGACCTGAAGCGGCTGAACCTACACCCAGCGCTCTGTACG	1036
PT	resistance, useful for conferring resistance or tolerance to a plant			QY	1064	CGTTCAGCGGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1123
PT	pathogen.			Db	1037	AGAACATGCGGCTGTTCCCGCGCGGTCGAGTTTCACTTCAAGTTCTGCGCGCGCGGAGC	1096
XX	Claim 1; SEQ ID NO 373; 299pp; English.			QY	1124	ACGCCAGACAAAGGCGTTTCGCGCGCGCGGTCGCGGAGGAGTTTCGCGGCGCTCTCACT	1183
XX	The invention relates to a novel isolated polynucleotide comprising a			Db	1097	TGCTCCCGAGCGCACCTACGTCGACGCGCGCGCGGCGCTCATGTACACCCCTACGCCA	1156
XX	plant nucleotide sequence having an open reading frame that encodes a			QY	1184	ACAGCTTCGGCAAGTGGTCTTACCTCCACGCTCGCTCAC	1223
CC	polypeptide associated with disease resistance or its fragment having			Db	1157	TGGCGCGCATGCGCGCATCTGCGGCGCGCGGAGCTGCGAGC	1196
CC	substantially the same activity as the full-length polypeptide. The			RESULT 12			
CC	polynucleotide of the invention is useful for conferring resistance or			ID	ABZ13076	standard; DNA; 1662 BP.	
CC	tolerance to a plant pathogen. The present sequence represents a gene			XX	ABZ13076		
CC	conferring disease resistance used in the invention.			AC	ABZ13076;		
XX	Sequence 1395 BP; 196 A; 497 C; 475 G; 227 T; 0 U; 0 Other;			XX	21-JAN-2003	(first entry)	
XX	Query Match 7.9%; Score 150.8; DB 8; Length 1395;			DT	Arabidopsis thaliana	stress regulated gene SEQ ID NO 881.	
XX	Best Local Similarity 49.2%; Pred. NO. 2.8e-19;			DE	Arabidopsis thaliana;	plant; gene; stress; transgenic; ds.	
XX	Matches 522; Conservative 0; Mismatches 517; Indels 21; Gaps 4;			XX	Arabidopsis thaliana.		
QY	176	CGGTGGACGTGAGGAACCTACCAACCGGATGACGAGTGTGCGGTGACCTGTGCTCAC	235	XX	WO200216655-A2.		
Db	146	CGGGTCTGTGAGGAGGAGTTGCGCAACCTCGGCGATGTGTACGCGGACCTGTCCGCC	205				
QY	236	GGCAGGACGAGTACCGGTGACATGCGGTTCACTTCTTACACCTACATTCGCTGACCCGG	295				
Db	206	GCTCCGCGAGGCGCCGCTCCACGTCCTCGGCTGACCTGCGGCTGACCGCGGAGCCGG	265				
QY	296	TGAATGTGAGATGCTTCAGACTACTTACCAATTTACCCCAAGGGAATCGGTACA	355				
Db	266	CGAAGTGGAGTACATGTCTCAAGCGCGCTTCGCAACTTCCCAAGGGAGCGGCTGCG	325				
QY	356	GATCCTACATGACGCTGCTCTCGGTGACGGATCTTCAACCGCGGAGCTGTGGA	415				
Db	326	CGCGGCTCTCTCGGCGACCTCTCTCGGCGAGGATCTTCAACGTCGACGCGCGGTGCG	385				

[illegible]

Db	724	GT	CAGCTT	GAGCGG	AGCCT	AGG	GAGAT	CGAT	TGGGTATTTAGATGCTGTCATTAA	TACA	783		
QY	808	AG	AGAGCG	CGAGAT	CGT	CAGGT	CCGGC	CAGCG	CAACAGG	AGAAATG	AGCACCGAC	867	
Db	784	CG	TAGC	ACAGAA	TTGCT	GAGT	CAGCG	GAGAG	TGG-----GGTCCAGCGT	CACGACGAT	837		
QY	868	AT	CTCTGT	CA	CGGTTC	ATC	GAGCT	GGGCG	HGGCGG	CGGACG	CGCGGCTTCGGGGAC	927	
Db	838	CT	CTCTCT	CTGTT	CAT	TGAAGA	AAAGA-----	-----	-----	CCAGTCTGTAC	AGC	879	
QY	928	GA	T	AGAGC	CTCGG	ACG	TGGTCT	CAACTT	CGTGAT	CGCCGGCGG	GACACGACGCGG	987	
Db	880	GAG	ACGTTCT	TACG	ACAGCT	GGCGCTT	AACTT	CACT	CTAGCT	GGHAGCT	GACACGTCATCA	939	
QY	988	AC	GACGCT	GTCTGT	GTTC	ACG	CAAT	GGCCAT	GTCCAC	CCCGGAC	GTGGCCGAGAGCTG	1047	
Db	940	GT	AGCGTT	GAGCT	TGGTTT	TCT	GGCTCAT	CAC	GACGAT	CTCCACGCTT	GAGGATAAGATC	999	
QY	1048	CG	CGCGAG	CTGTG	CGGCTT	CGA	GGCGG	ACGCGCG	CGCGCG	GAGGAGG	CGCTC	ACGCTCGTG	1107
Db	1000	GT	CGCGAG	ATATG	CTC-----	-----	-----	-----	-----	CGTCTGATTTAG	ACACGCTGGAAC	1040	
QY	1108	CT	CTGCGGCG	CGCTG	ACGCG	ACG	CAAGG	CGTTG	CGCCCGCCG	CTGGCGC	AGTTTCGCG	1167	
Db	1041	CG	ATGTA	TCGTC	GGAGC-----	-----	-----	-----	-----	-----	-----	1062	
QY	1168	GG	CTCTCT	CACTAC	GACAGC	CTT	CGG	CAAGCT	TGTTACT	TCCACG	CGCTCGTCA	CCGAG	1227
Db	1063	GAG	CGGTTG	GAATTC	GAT	GAGT	TCG	ACCGGTT	TGGTTT	TA	CTGAAGCGCG	CGCTCTCTGAG	1122
QY	1228	AC	GTC	CGCCTGT	ATACCC	CGCGCT	CCCTC	AGGACCC	CAAGGGAT	CTCTG	GAGGACG	AGCTG	1287
Db	1123	AC	GTTAGG	CTTTAC	CGCTG	CTCCG	GAATTC	CA	AAGACG	CTGCTGA	ACGAGATC	1182	
QY	1288	CT	GCGGAC	GGAGCG	AC	MA	GCTGAG	GGCGCGG	GGATGCT	GACGTAC	GTCCTACT	CGATG	1347
Db	1183	TT	ACCGGAC	GGAATTC	TTCG	TACCG	CGGAT	TCGTC	GGTGACT	TTATCG	ATCTAC	CGCGCG	1242
QY	1348	GG	CGGAT	TGAGTA	CAACT	GTGGG	CCCCG	ACGCG	CGAGCTT	CCGGCGG	AGCGGTG	GATC	1407
Db	1243	GG	AGGAT	TGA	GAGCAC	GTG	GGGAG	AGGATTC	TGGAATTC	CA	AACCGG	GAGGTGGATC	1302
QY	1408	---	AA	CGAGGAT	TGGCG	CTTCGG	CAACG	CGCTCG	CGGTTC	CAAGTT	TCAGGGCTT	CCAGCG	1454
Db	1303	TC	CGCGGAC	GTGA	TGGAAATTC	GT	GAT	CCGATTC	CGATTC	GGCGGCTTT	TAACGCC	1362	
QY	1465	GG	CGCGAG	ATCTG	CTGGG	CAAGG	ACTCG	CGCTAC	CTG	CAGATCA	AGATCG	CGCTGCC	1524
Db	1363	GG	ACCT	TAGGATCT	GTCTAG	GA	AAAGATCT	AGG	ATCTG	CAGAT	GTAG	ACGATG	1422
QY	1525	AT	CTCTTCC	CGCTT	CTC	AC	AGTTC	CGGTG	CTGTG	GAGGG	CGACCCG	GTCCAGT	1584
Db	1423	G	CGGTTT	ACTCAG	GCA	TAGATA	CT	ACCGT	TGGCG	CGGAC	CA	CAAGGT	1482
QY	1585	AT	GAC	ATCCTCT	CTC	AT	GCGCAC	CGGCT	CT	CAAGT	TC	CGGT	1625
Db	1483	TC	GTTC	ACATTTGTT	TCAT	GA	AGAAAC	CGGACTTT	TTGGT	CA	CGT	1523	

RESULT 13	
ACL21310/c	
ID	ACL21310 standard; DNA; 657 BP.
XX	
XX	
AC	ACL21310;
XX	
XX	27-OCT-2003 (revised)
DT	
DT	17-OCT-2003 (first entry)
XX	
XX	
DE	DNA clone originating in barley containing SNP encoding sequence #11301.
XX	
XX	Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KW	gene; ss.
KW	

Mon Mar 8 10:13:26 2004

XX OS Hordeum vulgare; var. (cul.Akashinriki).

XX PN WO2003057877-A1.

XX PD 17-JUL-2003.

XX EF 16-DEC-2002; 2002WO-IB005403.

XX PR 20-DEC-2001; 2001JP-00387059.

XX PR 20-DEC-2001; 2001JP-00387131.

XX PR 20-DEC-2001; 2001JP-00403299.

XX PR 20-DEC-2001; 2001JP-00403300.

XX PR 27-SEP-2002; 2002JP-00327515.

XX PA (UYN1-) UNIV JAPAN OKAYAMA.

XX PI Sato K, Takeda K, Kohara Y;

XX DR WPI; 2003-587127/55.

XX PT Single nucleotide polymorphism sites in barley varieties and DNA

PT sequences containing them for analysis and identification of barley

PT varieties and production of barley transformants with desired

PT characteristics.

XX PS Disclosure; SEQ ID XX; 284pp; Japanese.

XX CC The present invention relates to oligonucleotide clones originating in

CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms

CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley

CC varieties, identification of particular varieties and genotype-phenotype

CC analysis, isolation of specific genes and creation of new varieties by

CC transformation of barley varieties with them and production of new barley

CC varieties with desired properties. The present sequence represents an

CC oligonucleotide clone DNA sequence featured in the specification. The

CC single nucleotide polymorphism sites in barley varieties and DNA

CC sequences containing them for analysis and identification of barley

CC varieties and production of barley transformants with desired

CC characteristics.

XX SQ Sequence 657 BP; 140 A; 227 C; 159 G; 131 T; 0 U; 0 Other;

Query Match 7.6%; Score 144.6; DB 8; Length 657;

Best Local Similarity 59.3%; Pred. No. 3.9e-18;

Matches 265; Conservative 0; Mismatches 179; Indels 3; Gaps 1;

QY 1182 CGACAGCTCGGCAAGTGTCTACCTCCACGGCTGGTCCACGACGCTCCGCTGTA 1241

DB 651 CGACAGCTGAGGAGATGCACTACGTGACGCTGCCATCAGGAGTCAATGCACTTA 592

QY 1242 CCCGCGCTCCCTCAGGACCCCAAGGGGATCCTGGAGGACGAGTGTGCGGACGGAC 1301

DB 591 CCCGCGCTCCCTGAGTGAACCTGCTGACGGGAGGCGCGGAGCTTCTCCGACGGAC 532

QY 1302 GAAGGTGAGGCGCGCGGAGTGTGACGTAGTCCCTACTCGATGCGGCGGATGAGTA 1361

DB 531 GCGGTGGGACGGTGTGTTGCTGGGCGTCAACTCGTATGCAATGGGCGGATGGAGTC 472

QY 1362 CAATGGGGCCCCGACGCGGCGAGCTTCCGCGCGGAGGCTGGATCAAC---GAGGATGG 1418

DB 471 TGTGTGGGGCGAGGACCGGACGGGCTATCCGCGCGGAGCGGTGGACCGCGGAGGG 412

QY 1419 CGCGTTCGCAACCGCTCGCGCTTCAAGTTTCAACGGCGTTCCAGCGGGCGGACGATCTG 1478

DB 411 GACGTTCGCGCGGAGAGCCCGTTCGCGTATCGCGTATCGCGGTTTCCGCGGCGCAAAATTG 352

QY 1479 CCTGGCAAGGACTCGGCGTACCTGCAGATGAAGATGGCGCTGGCCATTCCTTCGCGTT 1538

DB 351 CCTCGAAAGGAGATGGCGTATATCCAGATGAAGTCTATCGTGGCGTGGTGAAGA 292

QY 1539 CTACAGCTTCGCGCTGTGGAGGGGACCCCGGTGACGATGACGATGACATCCCTCTC 1598

DB 291 GTTTGAATTGGCGTGGACGGCGGTACCGCGCGGCGAGGTGGCATCACTCAGCTTAAG 232

QY 1599 CATGGCGACGGCTCAAGGTCCGGT 1625

DB 231 AATGGCGGACGGGTCCCGGTGGGGT 205

RESULT 14

ID ACL21306/c

XX ACL21306 standard; DNA; 621 BP.

XX AC ACL21306;

XX DT 27-OCT-2003 (revised)

XX DT 17-OCT-2003 (first entry)

XX DE DNA clone originating in barley containing SNP encoding sequence #11297.

XX KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;

XX KW gene; ss.

XX OS Hordeum vulgare; var. (cul.Akashinriki).

XX PN WO2003057877-A1.

XX PD 17-JUL-2003.

XX PF 16-DEC-2002; 2002WO-IB005403.

XX PR 20-DEC-2001; 2001JP-00387059.

XX PR 20-DEC-2001; 2001JP-00387131.

XX PR 20-DEC-2001; 2001JP-00403299.

XX PR 20-DEC-2001; 2001JP-00403300.

XX PR 27-SEP-2002; 2002JP-00327515.

XX PA (UYN1-) UNIV JAPAN OKAYAMA.

XX PI Sato K, Takeda K, Kohara Y;

XX DR WPI; 2003-587127/55.

XX PT Single nucleotide polymorphism sites in barley varieties and DNA

PT sequences containing them for analysis and identification of barley

PT varieties and production of barley transformants with desired

PT characteristics.

XX PS Disclosure; SEQ ID XX; 284pp; Japanese.

XX CC The present invention relates to oligonucleotide clones originating in

CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms

CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley

CC varieties, identification of particular varieties and genotype-phenotype

CC analysis, isolation of specific genes and creation of new varieties by

CC transformation of barley varieties with them and production of new barley

CC varieties with desired properties. The present sequence represents an

CC oligonucleotide clone DNA sequence featured in the specification. The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to

CC standardise OS field)

XX SQ Sequence 621 BP; 111 A; 231 C; 180 G; 97 T; 0 U; 2 Other;

Query Match 7.2%; Score 137.2; DB 8; Length 621;

Best Local Similarity 65.3%; Pred. No. 9.7e-17;

Matches 218; Conservative 0; Mismatches 113; Indels 3; Gaps 1;

QY 1182 CGACAGCTCGGCAAGTGTCTACCTCCACGCTGGTCCACGACGCTCCGCTGTA 1241

DB 428 CGACAGCTGAGGAGATGACATGCTGACGCTGCCATCAGGATCATGCACTTA 369

QY 1242 CCCGCGCTCCCTCAGGACCCCAAGGGGATCTCTGGAGGACGAGTGTCTGCGGACGGAC 1301



368	CCGCGCGGTCCCGTGAAC	TCGCTGACGCGGAGCCGCGACGTTCTTCCGACGGCAC	309
	Ddb		
1302	GAAGGTGAGGCGCGCGGATGGT	GACGTACGTCCCTACTCGATGGGCGGATGGAGTA	1361
	Qy		
308	GGCGGTGGGACGGGTGGTT	CGTGCGTACAACTCGTATGCAATGGGCGGATGGAGTC	249
	Ddb		
1362	CAACTGGGCGCCGACGCGCGGAGCTT	CCGCGCGGAGCGGTGGATCAAC---GAGGATGG	1418
	Qy		
248	TGTGTGGGCGAGGACGCGCAGCGT	ATCCGGCCGAGCGGTGGCTGGACCCGCGGAGGG	189
	Ddb		
1419	CGCGTTCGCGCAACGCGTCGCGCTT	CAAGTTCACGGCGGTTCCAGGCGGGGCCGAGGATCTG	1478
	Qy		
188	GACGTTCCGCGCGGAGACCCGTT	CCGGTACATGGCGTTTCACGCGGGGCCAAGAAATTG	129
	Ddb		
1479	CTTGGGCAAGGACTCGCGGTACT	CGAGATGAAG	1512
	Qy		
128	CTCTCGAAAGGAGATGGCGTATAT	CCAGATGAAG	95
	Ddb		

RESULT 15

ABZ14873  
ID ABZ14873 standard: DNA: 1614 BP.

AC ABZ14873:

21-JAN-2003 (first entry)

XX  
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2678.

xx arabidonsis thaliana: njant: gene: stress: transgenic: ds:

XX Arabidopsis thaliana. OS

XX PN W0200216655-A2.

XX  
29-FEB-2002XX  
PF 24-AUG-2001. 2001WO-IIS026685.XX  
PB 24-2116-2000-2000US-0227866P

PR 26-JAN-2001; 2001US-0264647P.

PR 22-JUN-2001; 2001US-0300111P.  
yy

PA (SCRI ) SCRIPPS RES INST.

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX PI Harper JF, Kreps J, Wang X, Zhu T;

XX  
DE WPT. 3003-304137/34

XX Identifying a stress condition to which a plant cell has been exposed and  
 PT producing plants with increased tolerance to these abiotic stresses.

XX  
PS  
Claim 144: SEO ID NO 2678; 577pp + Sequence Listing; English.

The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising: (a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and (b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (AB212198-AB217574) used in methods of the invention. Note: The sequence data for the printed sequence is not represented in the printed specification but is based on sequence information supplied to, and received by, the European Patent Office

Sequence 1614 BP: 384 A; 365 C; 421 G; 444 T; 0 U; 0 Other; XX  
SO

Query Match	7.1%	Score 135.6;	DB 6;	Length 1614;
-------------	------	--------------	-------	--------------

Best Local Similarity	48.3%;	Pred. No. 2.2e-16;	Indels	102;
Matches	660.	Conservative	0:	Mismatches 604:

Figures 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1225, 1226, 1227, 1228, 1229, 1230, 1231, 1232, 1233, 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243, 1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253, 1254, 1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1286, 1287, 1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309, 1310, 1311, 1312, 1313, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348, 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400, 1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412, 1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448, 1449, 1450, 1451, 1452, 1453, 1454, 1455, 1456, 1457, 1458, 1459, 1460, 1461, 1462, 1463, 1464, 1465, 1466, 1467, 1468, 1469, 1470, 1471, 1472, 1473, 1474, 1475, 1476, 1477, 1478, 1479, 1480, 1481, 1482, 1483, 1484, 1485, 1486, 1487, 1488, 1489, 1490, 1491, 1492, 1493, 1494, 1495, 1496, 1497, 1498, 1499, 1500, 1501, 1502, 1503, 1504, 1505, 1506, 1507, 1508, 1509, 1510, 1511, 1512, 1513, 1514, 1515, 1516, 1517, 1518, 1519, 1520, 1521, 1522, 1523, 1524, 1525, 1526, 1527, 1528, 1529, 1530, 1531, 1532, 1533, 1534, 1535, 1536, 1537, 1538, 1539, 1540, 1541, 1542, 1543, 1544, 1545, 1546, 1547, 1548, 1549, 1550, 1551, 1552, 1553, 1554, 1555, 1556, 1557, 1558, 1559, 1560, 1561, 1562, 1563, 1564, 1565, 1566, 1567, 1568, 1569, 1570, 1571, 1572, 1573, 1574, 1575, 1576, 1577, 1578, 1579, 1580, 1581, 1582, 1583, 1584, 1585, 1586, 1587, 1588, 1589, 1590, 1591, 1592, 1593, 1594, 1595, 1596, 1597, 1598, 1599, 1600, 1601, 1602, 1603, 1604, 1605, 1606, 1607, 1608, 1609, 1610, 1611, 1612, 1613, 1614, 1615, 1616, 1617, 1618, 1619, 1620, 1621, 1622, 1623, 1624, 1625, 1626, 1627, 1628, 1629, 1630, 1631, 1632, 1633, 1634, 1635, 1636, 1637, 1638, 1639, 1640, 1641, 1642, 1643, 1644, 1645, 1646, 1647, 1648, 1649, 1650, 1651, 1652, 1653, 1654, 1655, 1656, 1657, 1658, 1659, 1660, 1661, 1662, 1663, 1664, 1665, 1666, 1667, 1668, 1669, 1670, 1671, 1672, 1673, 1674, 1675, 1676, 1677, 1678, 1679, 1680, 1681,

276	QY	CACCTCATCGCTGACCGGTAATGTCGAGCATGTCCTCAAGACTAACTTCCACCAATTA	335
243	DB	CACGTGCATCTGGCGATCAAGAACCCTCGAGCATATCTCAAGNATCGTTTGTATAATTA	302
336	QY	CCCCAAGGAATCGTGTACAGATCCCTACATGGACGTCTCTCGTGTGACGGATCTTCAA	395
303	DB	CCCTAAAGTCCAAACGTGGCAAGCGCTTTTCCATGATCTTCTCGGCCAAGGTATCTTTAA	362
396	QY	CGCCGACGGCAGCTGTGGAGAACGAGAGACAGCGCGAGTTCGAGTTCGCTCCAA	455
363	DB	CTCCGACGGTGACACGTGGCTTTTTCAGCGTGAACCCCGCTTGGAGTTCACCAAG	422
456	QY	GAACCTGAGGGATTTACGGCCATT---GTGTTCCAGAGATPACTCCCTGAAGCTGTCCGG	512
423	DB	GACACTCCGTCAAGCATTGGCTCGTGGGTAAATAGAGCGATCAAACTTAGGTTCTCGCC	482
513	QY	TATATCGACGACGGCATCCAGGACGAGCAAGTTGTGGACATGACAGAACTTTACATGAG	572
483	DB	TATCTTGAAATGCTCGGCTCGGCTCTGAGCCGATTGATCTTCAAGATTTCGTTCTTCG	542
573	QY	GATGAGCTGGACTCCATCTCAAGTTGGGTTCGGGTCGAGATCGGACAGCTGTCCGCC	632
543	DB	GCTTACATTTGATAATATCTCGGCTTAACTTTTCGGGAAAGACCCCGGACTTGTGCACC	602
633	QY	AGATCTCCCGGAGAACAGCTTCGGCAGCGGTTTGATCCGCCAACATCATCATCAGCT	692
603	DB	GGGTCTACCGGTGAACACATTCGCGGTGGCTTTTCGATCGAGCCACGGAGGCTTCGCTACA	662
693	QY	GGGGTTTCATCGACCG------CTGTGGCGCATCAAGAGGTTCTTCCACGTTCGGTCA	746
663	DB	AAGGTTTATATTTCGCGGAGATATTATGAAAGTTCAAGAGATGGCTCCGACTCGGCTTGA	722
747	QY	GCCTCTCTAGCGCAGAGCATCAAGCTCGTGGACAGGTTTCACCTACAGCGTGATCCGCGC	806
723	DB	AGTCAGCTTAAACAAGGCTTGGTCCAGGTAGATAATTTATTGCTGAGATTAATACGAC	782
807	QY	GAGGAAGGCGGAGATGTCGAGGTTCGGGCCAGCGGCGAAAACGAGAGAGATGAGCAGCA	866
783	DB	ACGTAAGGAAGAAATGAT-----GACTCAGCATATAATATGGAAACACCCACGACGA	833
867	QY	CATCTGTCTACGGTTCATCGAGCTGGCGAGCGCGGACGACGCGCGGCTTCGGGGA	926
834	DB	TCCTTTATCGCGGTTTCAAGAAAGAGGAGTCTTACTCCGAGC-----	877
927	QY	CGATAAGAGCCTCCGGGACGTGTGTCTCAACTTCGTGATTCGCGGCGGCGGAACGACGCG	986
878	DB	-----AGACCTCCAGCGGTGGCGCTCAACTTCATCTCTAGCTGGACGTGACACGTCA	932
987	QY	GACGAGCTGTGTGTTTACGCCATATGSCCATGTCCACCCGAGCTGGCGGAGAGCT	1046
933	DB	AGTCGGGCTGAGCTGTTCTTCTGTGTGATTAACGACGATCCGGCCATAGACGATATAAT	992
1047	QY	GCGCCGAGCTGTGCGGTTTCGAGCGGAGGCGCGCGCGAGGAGGGCGTCAACGCTGT	1106
993	DB	CCTCCGTGAGATCTGTACCGTTCTGTGTGAGACACGTGGCGATGATGTCG------	1043
1107	QY	GCTCTGGCGGCGGCTGACGCGACGACAGGGGTTTCGCGCCCGCTGGCGGATTCG	1166
1044	DB	-----GTTATGGACGGA	1055
1167	QY	GGGCTCTCAACCTACAGAGCTTCGGCAGCTGTGTCTACCTCCAGCGCTGCTCACCGA	1226
1056	DB	CGAGCCGCTTTCGTGGCAGAGCTTGATCGATTGGTTTTTCTCAGGCGAGCTTATCGGA	1115
1227	QY	GACGCTCGGCTGTACCCCGCGTCCCTCAGGACCCCAAGGGGATCCTCGAGGACGACGT	1286
1116	DB	GACTTTGAGGCTTTATCTCTCGGTGCGGAGGATTCAAAACGCGCGCTGAAGACGATGT	1175
1287	QY	GCTGCCGACGGGACGAAGGTGAGGGCCGCGGGAGTGTGACGTACGTGCCCTACTCGAT	1346
1176	DB	GTATCCGACGGGACATTTGTGCGCGGGGATCTTCGATTACTTATTCGATTACTCTGC	1235



us-10-021-657-1.rng

Mon Mar 8 10:13:26 2004

QY	1347	GGGGGGATGGAGTACAACTGGGGCCCGACGGCGGAGCTTCGGCGCGGAGCGGTGGAT	1406
Db	1236	GGGGAGGATGAAATCGACTTGGGAGAGGATTTTGGAAATCAACACAGAGATGGAT	1295
QY	1407	CAAC-----GAGGATGGCGGTTCCGCAACCGCGTCGCGTTCAAGTTACGGCGTTCCA	1460
Db	1296	TTGGCAATCGGACGGTGGAGATTCAATCATGATCCGTTTAAATTCGTTGCGTTCAA	1355
QY	1461	GGGGGGCCGAGGATCTGCTGGGCAAGGACTCGGCGTACCTGCAGATGAAGATGGCGCT	1520
Db	1356	TGCTGGCCCTAGGATCTCTGCTGAGGATCTGGCGTATCTACAAATGAAATCAATTGC	1415
QY	1521	GGCCATCCTCTTCCGCTTCTACAGCTTCGGGCTGCTGGAGGGGCACCGGTCAGTACCG	1580
Db	1416	ATCGGGGTTTTTGCTCCGGCACCGTTGACGGTGGTGACGGGGCATAAGTTGGAACAGAA	1475
QY	1581	CATGATGACCAATCCTCTCCATGGCGACGGCCTCAAGTCCGCGTC	1626
Db	1476	GATGTCGTTAACTTTATTTCATGAAGTACGGTCTTTTAGTTAACGTC	1521

Search completed: March 5, 2004, 22:48:59  
Job time : 815 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2004, 17:06:39 ; Search time 7438 Seconds  
(without alignments)  
11106.725 Million cell updates/sec

Title: US-10-021-657-1  
Perfect score: 1906  
Sequence: 1 gaattcggaagggaagc.....aaaaaaaaaaactcgag 1906

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*\*

- 1: gb\_ba:\*\*
- 2: gb\_htg:\*\*
- 3: gb\_in:\*\*
- 4: gb\_om:\*\*
- 5: gb\_ov:\*\*
- 6: gb\_pat:\*\*
- 7: gb\_ph:\*\*
- 8: gb\_pl:\*\*
- 9: gb\_pr:\*\*
- 10: gb\_ro:\*\*
- 11: gb\_sts:\*\*
- 12: gb\_sy:\*\*
- 13: gb\_un:\*\*
- 14: gb\_vl:\*\*
- 15: em\_ba:\*\*
- 16: em\_fun:\*\*
- 17: em\_hum:\*\*
- 18: em\_in:\*\*
- 19: em\_mu:\*\*
- 20: em\_om:\*\*
- 21: em\_or:\*\*
- 22: em\_ov:\*\*
- 23: em\_pat:\*\*
- 24: em\_ph:\*\*
- 25: em\_pl:\*\*
- 26: em\_ro:\*\*
- 27: em\_sts:\*\*
- 28: em\_un:\*\*
- 29: em\_vl:\*\*
- 30: em\_htg\_hum:\*\*
- 31: em\_htg\_inv:\*\*
- 32: em\_htg\_other:\*\*
- 33: em\_htg\_mus:\*\*
- 34: em\_htg\_pln:\*\*
- 35: em\_htg\_rod:\*\*
- 36: em\_htg\_mam:\*\*
- 37: em\_htg\_vrt:\*\*
- 38: em\_sy:\*\*
- 39: em\_htgo\_hum:\*\*
- 40: em\_htgo\_mus:\*\*
- 41: em\_htgo\_other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1906	100.0	1906	6	AX453572	AX453572 Sequence
2	1886.2	99.0	1891	8	AF366297	Zea mays
3	1378.4	72.3	3897	6	AX453578	AX453578 Sequence
4	861.4	45.2	160738	8	AC073556	AC073556 Oryza sat
5	548.8	28.8	1695	8	AK118803	AK118803 Arabidops
6	433.2	22.7	1494	6	AX453574	AX453574 Sequence
7	366.6	19.2	1714	8	AK119612	AK119612 Oryza sat
8	366.6	19.2	1780	8	AK103182	AK103182 Oryza sat
9	366.6	19.2	1904	8	AK101618	AK101618 Oryza sat
10	363.4	19.1	1712	8	AK102091	AK102091 Oryza sat
11	361.4	19.0	1934	8	AK099925	AK099925 Oryza sat
12	361.4	19.0	1934	8	AK099925	AK099925 Oryza sat
13	359.8	18.9	1916	8	AK104739	AK104739 Oryza sat
14	293.8	15.4	1881	8	AK061650	AK061650 Oryza sat
15	290.2	15.2	2118	8	AK069029	AK069029 Oryza sat
16	290.2	15.2	110492	2	AP004139	AP004139 Oryza sat
17	290.2	15.2	154138	2	AP005108	AP005108 Oryza sat
18	272.8	14.3	2095	8	AK106001	AK106001 Oryza sat
19	272.8	14.3	147806	8	OSJNQ0087	OSJNQ0087 Sequence
20	271.6	14.2	1626	6	AR126779	AR126779 Sequence
21	271.6	14.2	1626	6	AX023070	AX023070 Sequence
22	271.6	14.2	1626	6	AX031115	AX031115 Sequence
23	271.6	14.2	2181	6	AR126775	AR126775 Sequence
24	271.6	14.2	2181	6	AX023066	AX023066 Sequence
25	271.6	14.2	2181	6	AX031111	AX031111 Sequence
26	271.6	14.2	2181	6	AF123609	AF123609 Triticum
27	268.4	14.1	1935	8	AK107532	AK107532 Oryza sat
28	268.4	14.1	134821	8	AC137697	AC137697 Genomic S
29	268.4	14.1	156647	8	AC137930	AC137930 Oryza sat
30	265	13.9	1371	8	AK105884	AK105884 Oryza sat
31	251.2	13.2	146519	8	CNS08C9V	AL772417 Oryza sat
32	251.2	13.2	174521	8	CNS08C8D	AL731873 Oryza sat
33	246	12.9	148816	8	AC087182	AC087182 Oryza sat
34	246	12.9	284663	8	AB017107	AB017107 Oryza sat
35	239.2	12.5	150053	8	AP003735	AP003735 Oryza sat
36	232.6	12.2	1564	8	AK102714	AK102714 Oryza sat
37	232.6	12.2	165497	8	AP003232	AP003232 Oryza sat
38	228	12.0	2038	8	AK109676	AK109676 Oryza sat
39	228	12.0	150976	8	AP003442	AP003442 Oryza sat
40	226.4	11.9	1527	6	AX654049	AX654049 Sequence
41	225.4	11.8	1757	8	AK070931	AK070931 Oryza sat
42	220.6	11.6	1658	8	AK119195	AK119195 Oryza sat
43	219	11.5	1549	8	AK067844	AK067844 Oryza sat
44	216	11.3	1471	8	AK071033	AK071033 Oryza sat
45	214.4	11.2	1563	6	AX654116	AX654116 Sequence

ALIGNMENTS

RESULT 1	AX453572	1906 bp	DNA	linear	FAT 02-SEP-2002
LOCUS	Sequence 1 from Patent WO0226789.				
DEFINITION	AX453572				
VERSION	AX453572.1	GI:21712815			
KEYWORDS	Zea mays				
SOURCE	Zea mays				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD				
	clade; Panicoideae; Andropogoneae; Zea.				
REFERENCE	1				
AUTHORS	Albertsen,M.C., Fox,T., Huffman,G. and Trimmell,M.				
TITLE	Nucleotide sequences mediating male fertility and method of using				

same		100.0%; Score 1906; DB 6; Length 1906;	
JOURNAL	Patent: WO 0226789-A 1 04-APR-2002;	Query Match	100.0%; DB 6; Length 1906;
FEATURES	PIONEER HI-BRED INTERNATIONAL, INC. (US)	Best Local Similarity	100.0%; Pred. No. 2.8e-228;
source	Location/Qualifiers	Matches 1906; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
CDS	1..1906	1	GAATTCGGACAGGAGGAGCTACCTCAGCGCGGAGCGCCATCGGCCATTCCTCCACTA 60
	/organism="Zea mays"	1	GAATTCGGACAGGAGGAGCTACCTCAGCGCGGAGCGCCATCGGCCATTCCTCCACTA 60
CDS	/mol_type="unassigned DNA"	61	GCAGGGCTCACAGTACATCGCGCTCTTCCTGCTGCTCTCATGAGTCTGCTGCTG 120
	/db_xref="taxon:4577"	61	GCAGGGCTCACAGTACATCGCGCTCTTCCTGCTGCTCTCATGAGTCTGCTGCTG 120
CDS	join(<1..1638,1642..1770)	121	AGTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
	/note="unnamed protein product"	121	AGTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
CDS	/codon_start=1	181	GAGCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
	/protein_id="CAD38131.1"	181	GAGCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
CDS	/db_xref="GI:21712816"	241	AGGACAGTACCGTGCAGTCCGCTTCACTTCTTCACTTCACTTCACTTCACTTCA 300
	/db_xref="REMBL:CAD38131"	241	AGGACAGTACCGTGCAGTCCGCTTCACTTCTTCACTTCACTTCACTTCACTTCA 300
CDS	translation="EFGTREAHLPATPSPPFLAGPHKXIALLVLSWILVORWSL RKQKPRNFWPVGATVEOLRNVHRMDLGVLSRRTVTVDMPFTSYIADPNVVE HVLKTFNYPAGIVYRSIMVLLDGGIENADGELWRKQRTASFEFASKNLRDPSAI VEREYSLKSLGSLASQAGKGVDMQELVIMRMLDSICKVGFVIGILSPDLPSNF AQAFDAANIIITRFDPLMRKFRFHVHSEALLAOSIKLVDEFYIVIRKABIVE VRASGQKMKHDIILSRFELGEAGDGGGDKSLRDLVNLNFVIAGRDFTATLSW FTMHMSKEDVAEKRLRELCAPEAREEVEGTVLVCGGADADDFAARVAQFAGLL TVDSLGKLVILHACTETLRLYPVQPKGILEDDLPDGTFKVRAGMVTVPYSMG RMEYNWPDPAASFRERWINEGDAFRNASPFKPTAFQAGPRICLCKDSAYLQMKWALA ILFRYSFELLBEGHPVQYHNMVITLSMAHLKVRVSRVAVCHGDLMDIVPLNPRQITLV LQICMHACKGRWVSLVAVLKP"	301	GTGAGCATGTCCTCAAGACTAATCTTCAACAAATACCCCAAGGGAATCGTGACAGATCC 360
		301	GTGAGCATGTCCTCAAGACTAATCTTCAACAAATACCCCAAGGGAATCGTGACAGATCC 360
CDS		361	TACATGGAAGCGTCTCTCGGTGACGGGATCTTCAACCGCGAGCGGAGGAGGAGGAG 420
		361	TACATGGAAGCGTCTCTCGGTGACGGGATCTTCAACCGCGAGCGGAGGAGGAGGAG 420
CDS		421	CAGAGGAGAGCGGAGGAGTTCAGAGTTCGCTTCCAGAACCTGAGGAGTTTCAGGCGCAT 480
		421	CAGAGGAGAGCGGAGGAGTTCAGAGTTCGCTTCCAGAACCTGAGGAGTTTCAGGCGCAT 480
CDS		481	GTGTTTCAAGAGTACTCTCCCTGAAGCTGTCGGGTATCTAGCCAGGAGGATCCAAAGGAGG 540
		481	GTGTTTCAAGAGTACTCTCCCTGAAGCTGTCGGGTATCTAGCCAGGAGGATCCAAAGGAGG 540
CDS		541	AAAGTTGTGGAGCATGAGGAATTTTACATGAGGATGACGCTGGACTCCATCTGCAAGGTT 600
		541	AAAGTTGTGGAGCATGAGGAATTTTACATGAGGATGACGCTGGACTCCATCTGCAAGGTT 600
CDS		601	GGGTTCCGGGTCAGATCGGACGCTGTCGCGAGATCTCCCGGAGAACAGCTTCGCGCAG 660
		601	GGGTTCCGGGTCAGATCGGACGCTGTCGCGAGATCTCCCGGAGAACAGCTTCGCGCAG 660



```
1270 ATCTTGAGGACGACGCTGCTCGGACGGGACGAAGGTGAGGCGCGGGATGTTGACG 1329
Db ATCTTGAGGACGACGCTGCTCGGACGGGACGAAGGTGAGGCGCGGGATGTTGACG 1320
QY TAGTGGCCCTACTCGATGGGGCGGATGGAGTAACTGGGGGCGCGGAGGTTTC 1389
Db TAGTGGCCCTACTCGATGGGGCGGATGGAGTAACTGGGGGCGCGGAGGTTTC 1380
QY CGGCGGAGCGGTGGATCAACGAGGATGGCGTTCGCAACGCTGCGCCGTTCAAGTTC 1449
Db CGGCGGAGCGGTGGATCAACGAGGATGGCGTTCGCAACGCTGCGCCGTTCAAGTTC 1440
QY ACGGCGTTCAGGCGGGCGGAGGATCTGCTCGGCGAAGGACTCGGCGTACCTGCGAGATG 1509
Db ACGGCGTTCAGGCGGGCGGAGGATCTGCTCGGCGAAGGACTCGGCGTACCTGCGAGATG 1500
QY AAGATGGCGTGGCGCATCTCTCCATGCGCGCACGGCTCAAGGTCCGCGTCTCT 1629
Db AAGATGGCGTGGCGCATCTCTCCATGCGCGCACGGCTCAAGGTCCGCGTCTCT 1620
QY GTGCAGTACCGCATGATGACCATCTCTCCATGCGCGCACGGCTCAAGGTCCGCGTCTCT 1629
Db GTGCAGTACCGCATGATGACCATCTCTCCATGCGCGCACGGCTCAAGGTCCGCGTCTCT 1620
QY AGGCGCGTCTGATGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATG 1689
Db AGGCGCGTCTGATGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATG 1680
QY ATACGCGTCTGTTACAAATTTGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATG 1749
Db ATACGCGTCTGTTACAAATTTGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATG 1740
QY GTGCGTCTGCGTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1809
Db GTGCGTCTGCGTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1800
QY TTTTGTACTACTCTCTCAGTGGGAGTGTAGTACAGATGCGGATGCGGATGCGGATGCGGAT 1869
Db TTTTGTACTACTCTCTCAGTGGGAGTGTAGTACAGATGCGGATGCGGATGCGGATGCGGAT 1860
QY ATCCTCTCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1900
Db ATCCTCTCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1891

RESULT 3
AX453578
LOCUS AX453578 3897 bp DNA linear PAT 02-SEP-2002
DEFINITION Sequence 7 from Patent WO0226789.
ACCESSION AX453578
VERSION AX453578.1 GI:21712820
KEYWORDS
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1
AUTHORS Albertsen, M.C., Fox, T., Huffman, G. and Trimmell, M.
TITLE Nucleotide sequences mediating male fertility and method of using
same
JOURNAL Patent: WO 0226789-A 7 04-APR-2002;
PIONEER HI-BRED INTERNATIONAL, INC. (US)
FEATURES
source
1. 3897
/organism="Zea mays"
/mol_type="unassigned DNA"
/db_xref="taxon:4577"
ORIGIN
Query Match 72.3%; Score 1378.4; DB 6; Length 3897;
Best Local Similarity 82.1%; Pred. No. 1.2e-162;
```

```
Matches 1833; Conservative 0; Mismatches 16; Indels 385; Gaps 6;
QY 9 CACGAGGGAAGCTCACCCTCAGCGCGGACGCGCATCGCCATTCTTCCCACTAGCAGGCG 68
Db 1088 CATGGAGGAAGCTCACCCTCAGCGCGGACGCGCATCGCCATTCTTCCCACTAGCAGGCG 1147
QY 69 TCACAAGTACATCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 128
Db 1148 TCACAAGTACATCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1207
QY 129 CTTGAGGAAGCAGAAAGCGCGGATCATGGCCAGTCACTCGCGCAACCGTGGAGCAGCT 188
Db 1208 CTTGAGGAAGCAGAAAGCGCGGATCATGGCCAGTCACTCGCGCAACCGTGGAGCAGCT 1267
QY 189 GAGGAATACCAACCGGATGACGACTGCTTGTGCGGTACCTGTCAAGGCAACGAGCAGT 248
Db 1268 GAGGAATACCAACCGGATGACGACTGCTTGTGCGGTACCTGTCAAGGCAACGAGCAGT 1327
QY 249 GACCGTGCATGCGCTTCACTTCTTCACTTCTTCACTTCTTCACTTCTTCACTTCTTCACT 308
Db 1328 GACCGTGCATGCGCTTCACTTCTTCACTTCTTCACTTCTTCACTTCTTCACTTCTTCACT 1387
QY 309 TGTCTCTCAAGACTAACTTCAACCAATTAACCCCAAG----- 342
Db 1388 TGTCTCTCAAGACTAACTTCAACCAATTAACCCCAAGGTAATGACCTGAACCTCACTGATGT 1447
QY 343 -----GGAA 346
Db 1448 CAGTCTTCGAAATCAGAGCTGAAAGCTGAATGAACTGTGCTTGAACACCGTGTAGGGAA 1507
QY 347 TCGTGTACAGATCTTACATGAGCGTCTCTCTCGGTGACGCGCATCTTCAACGCGGACGCG 406
Db 1508 TCGTGTACAGATCTTACATGAGCGTCTCTCTCGGTGACGCGCATCTTCAACGCGGACGCG 1567
QY 407 AGCTGTGAGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 466
Db 1568 AGCTGTGAGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1627
QY 467 ATTTAGCGCATGTTGTTCAAGAGATCTCCCTGAAGCTGTGCGGTATCTGAGCAGG 526
Db 1628 ATTTAGCGCATGTTGTTCAAGAGATCTCCCTGAAGCTGTGCGGTATCTGAGCAGG 1687
QY 527 CATCCAAAGCAGCAAGTGTGACATG----- 555
Db 1688 CATCCAAAGCAGCAAGTGTGACATGAGTGTGACATGAGTGTGACATGAGTGTGACATG 1747
QY 556 -----CAGGAACCTTTACATG 570
Db 1748 AACATGAGCATTTTCAACCTTGAGACAGAGAGCTTACCTTGCAGATTTCAGGAACCTTTACATG 1807
QY 571 AGGATGAGCGTGGACTCCATCTGCAAGGTTGGTTGCGGGTGGAGATCGGACGCTGTGCG 630
Db 1808 AGGATGAGCGTGGACTCCATCTGCAAGGTTGGTTGCGGGTGGAGATCGGACGCTGTGCG 1867
QY 631 CCAGATCTCCCGCAGAACAGCTTCGCGCAGGCGTTCGATGCGGCCCAACATCATCATCACG 690
Db 1868 CCAGATCTCCCGCAGAACAGCTTCGCGCAGGCGTTCGATGCGGCCCAACATCATCATCACG 1927
QY 691 CTGCGGTTTCACTGACCCGCTGTGCGGCATCAAGAGGTTCTTCCACGTCGGGTGAGAGGCC 750
Db 1928 CTGCGGTTTCACTGACCCGCTGTGCGGCATCAAGAGGTTCTTCCACGTCGGGTGAGAGGCC 1987
QY 751 CTCTTAGCGCAGAGATCAAGCTGCTGGAGAGTTCACCTAGAGCTGATCCGCGGAGG 810
Db 1988 CTCTTAGCGCAGAGATCAAGCTGCTGGAGAGTTCACCTAGAGCTGATCCGCGGAGG 2047
QY 811 AAGCGGAGATCTGCGAGTCCGCGCAGCGGCAAAACAGGAGA----- 853
Db 2048 AAGCGGAGATCTGCGAGTCCGCGCAGCGGCAAAACAGGAGAAGGTACGTGCACATGAC 2107
QY 854 ----- 853
Db 2108 TGTTCGATTTCTCAGTTTCATCGTCTTGGCGGGATGACCTGTGATCTCTGATTATTATAT 2167
```

QY 854 -----AGATGAGCAGCAGATCTGTGACG 878  
Db 2168 ATCCGTGTGACTGTGAGGACAAATTAATGAGCAGATGAAGCAGCAGATCTGTGACG 2227  
QY 879 GTTCATCGAGCTGGCGGAGCGCGGAGCGGCGGCTTCGGGAGCAGATAGAGCCT 938  
Db 2228 GTTCATCGAGCTGGCGGAGCGCGGAGCGGCGGCTTCGGGAGCAGATAGAGCCT 2287  
QY 939 CCGGAGCGTGTGCTCAACTTCGTGATCGCGCGGCGGAGCAGCAGCGGCGGAGCAGCCTGTC 998  
Db 2288 CCGGAGCGTGTGCTCAACTTCGTGATCGCGCGGCGGAGCAGCAGCGGCGGAGCAGCCTGTC 2347  
QY 999 GTGGTTCACGACATGCGCATGTCCACCCGGAAGCTGCGGAGAGCTGCGCGGAGCT 1058  
Db 2348 GTGGTTCACGACATGCGCATGTCCACCCGGAAGCTGCGGAGAGCTGCGCGGAGCT 2407  
QY 1059 GTGGCGTTCGAGCGGAGCGCGGAGGAGGCTCAGCTCGTGTGCTGCTCGGCGG 1118  
Db 2408 GTGGCGTTCGAGCGGAGCGCGGAGGAGGCTCAGCTCGTGTGCTGCTCGGCGG 2467  
QY 1119 CGTGACGCGGAGCAGCAGGCGGTTGCGCGCGGCGGAGGCTGCGGCGGCTCCTCAC 1178  
Db 2468 CGTGACGCGGAGCAGCAGGCGGTTGCGCGCGGCGGAGGCTGCGGCGGCTCCTCAC 2527  
QY 1179 CTAGCAGAGCTCGGAGAGCTGTCTACCTCGAGCGCTGCTGCTGCTGCTGCTGCT 1238  
Db 2528 CTAGCAGAGCTCGGAGAGCTGTCTACCTCGAGCGCTGCTGCTGCTGCTGCTGCT 2587  
QY 1239 GTACCCGCGCTCCCT----- 1254  
Db 2588 GTACCCGCGCTCCCTCAGGTGAGCGCGCGGAGCAGCGGAGCTCGGTCGAGCAGCAGC 2647  
QY 1255 -----CAGGACCCCAAGG 1267  
Db 2648 ATGCAGTGAGTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2707  
QY 1268 GATCTCTGAGGAGCAGCTGTGCGGAGCGGAGCAGGAGTGAAGTGAAGTGAAGTGA 1327  
Db 2708 GATCTCTGAGGAGCAGCTGTGCGGAGCGGAGCAGGAGTGAAGTGAAGTGAAGTGA 2767  
QY 1328 CGTAGCTCCCTACTCGATGCGGCGGATGGAGTCAACTGGGCGCGGCGGCGGAGCT 1387  
Db 2768 CGTAGCTCCCTACTCGATGCGGCGGATGGAGTCAACTGGGCGCGGCGGCGGAGCT 2827  
QY 1388 TCCGCGCGGAGCGGTGATCAACGAGGATGGCGGCTTCGCAACGCGTCCGCTTCAAGT 1447  
Db 2828 TCCGCGCGGAGCGGTGATCAACGAGGATGGCGGCTTCGCAACGCGTCCGCTTCAAGT 2887  
QY 1448 TCAGCGGCTTCAGCGGCGGAGGATGCTGCTGGGCAAGGACTCGGCTACCTGCGAGA 1507  
Db 2888 TCAGCGGCTTCAGCGGCGGAGGATGCTGCTGGGCAAGGACTCGGCTACCTGCGAGA 2947  
QY 1508 TGAAGATGCGGCTGCGGCTTCCTTCGCTTCAGCTTCGCGGCTGCTGGAGGCGGAC 1567  
Db 2948 TGAAGATGCGGCTGCGGCTTCCTTCGCTTCAGCTTCGCGGCTGCTGGAGGCGGAC 3007  
QY 1568 CGGTGAGTACCGCATGATGACCATCTCTCCATGCGGCGGCTTCAGAGTTCGCGCT 1627  
Db 3008 CGGTGAGTACCGCATGATGACCATCTCTCCATGCGGCGGCTTCAGAGTTCGCGCT 3067  
QY 1628 CTAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1687  
Db 3068 CTAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3117  
QY 1688 AAATAACGCTCGTGTACAAATTCGATGATGATGATGATGATGATGATGATGATGAT 1747  
Db 3118 -----TTAAATAATTCGATGATGATGATGATGATGATGATGATGATGATGAT 3163  
QY 1748 TGGTGGCTTGGCTTAAAGCCTTAAAGCCTTAAAGCCTTAAAGCCTTAAAGCCTTAAAG 1807  
Db 3164 TGGTGGCTTGGCTTAAAGCCTTAAAGCCTTAAAGCCTTAAAGCCTTAAAGCCTTAAAG 3223

QY 1808 TGTCTTGTACTCTACTCTCCTCAGTGGAGTGTAGTGCAGCATCAAGTTCATCATATATA 1867  
Db 3224 TGTCTTGTACTCTACTCTCCTCAGTGGAGTGTAGTGCAGCATCAAGTTCATCATATATA 3283  
QY 1868 TTATCTCTCTTCTT 1881  
Db 3284 TTATCTCTCTTCTT 3297  
RESULT 4  
AC073556 160738 bp DNA linear PLN 24-APR-2002  
LOCUS Oryza sativa chromosome 3 BAC OSJNBa0091P11 genomic sequence,  
DEFINITION complete sequence.  
AC073556  
VERSION AC073556.11 GI:18997259  
KEYWORDS HTG.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Rhizophytidae; Oryzaceae; Oryza.  
REFERENCE 1 (bases 1 to 160738)  
AUTHORS Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Moffat, K.S., Hill, J.N.,  
Gansberger, K., Brenner, M., Burgess, S., Hance, M., Shvartsbeyn, M.,  
Tsitrin, T., Riggs, F., Hsiao, J., Zismann, V., Blunt, S., Pai, G.,  
VanAken, S.E., Utterback, R.R., Feldblyum, T.V., Kalb, E.,  
Quackenbush, J., Salzberg, S.L., White, O., and Fraser, C.M.  
Oryza sativa chromosome 3 BAC OSJNBa0091P11 genomic sequence  
Unpublished  
2 (bases 1 to 160738)  
AUTHORS Buell, R.  
REFERENCE Direct Submission  
TITLE Submitted (23-JUN-2000) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA  
AUTHORS 3 (bases 1 to 160738)  
REFERENCE Direct Submission  
TITLE Submitted (26-FEB-2002) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA  
AUTHORS Buell, R.  
REFERENCE Direct Submission  
TITLE Submitted (28-FEB-2002) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA  
AUTHORS Buell, R.  
REFERENCE Direct Submission  
TITLE Submitted (02-MAR-2002) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA  
AUTHORS Buell, R.  
REFERENCE Direct Submission  
TITLE Submitted (24-APR-2002) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org  
COMMENT On Feb 28, 2002 this sequence version replaced gi:18921330.  
Addresses all correspondence to:rice@tigr.org

BAC clone OSJNBa0091P11 is from Oryza sativa chromosome 3  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.  
Genes were identified by a combination of several methods: Gene  
prediction programs including Fgenesh (<http://www.softberry.com/>),  
GENSCAN and GENSCAN+ (Chris Burge,  
<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky,  
<http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer  
(Mihaila Perlea and Steven Salzberg, contact mperlea@tigr.org),  
searches of the complete sequence against a peptide database and  
the plant EST database at RIGR (<http://www.tigr.org/cdb/tgi.shtml>).  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as unknown proteins.

Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Eddy, 1996). Genes encoding rRNAs are predicted by rRNAscan-SE (Smit, 1996). Simple repeats are identified by repeatmasker (Arian Smith, 1996).  
<http://genome.wustl.edu/eddy/tRNAscan-SE/>.  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## FEATURES

sources

1. .160738

/organism="Oryza sativa (japonica cultivar-group)"

/mol type="genomic DNA"

/cultivar="Nipponbare"

```
/db xref="taxon:39947"
```

```
/chromosome="3"
```

```
/map="S13095"
```

```
/map= 21303  
/clone="OSJNBa0091P11"
```

/note="japonica cultivar-group"

1. 2006

```

T. .2006
/aene="O%

```

/gene="CB008003".  
/note="near identical to sucrose transporter (c terminus)  
GB:BA24071 G:2723471 (*Oryza sativa*) (Plant Cell Physiol  
38 (12), 1389-1396 (1997)); EST C27284, D40515, AU08830  
from this gene"

FROM CLITS gene			
join (<)	47	267	467
	580	723	833
	1066	1169	1232



QY 107 GGATCTCTGGTCAGAGGTGGACCTGAGGAAGCAGAAAGCCCGGAGATCATGGCCAGTCA 166  
DB 95750 GGATCTTGGTCACCAAGTGGAGCCTGAGGAACACGAAAGGGGCAAGATCATGGCCAAATCA 95809  
QY 167 TCGGCGCAACGGTGGAGCAGTGGAGAACTACCAACGGATGACAGACTGGCTTGTGCGGT 226  
DB 95810 TCGGCGCAACGGTGGAGCAGTGGAGAACTACCAACGGATGACAGACTGGCTTGTGCGGT 95869  
QY 227 ACCTGTACGGCACAAGGACAGTGAACGGTGCAGATGCGGTTCACCTTCTTACACCTACATG 286  
DB 95870 ACTTGTGAAGACAGGACGGTGAACGGTGCAGATGCGGTTCACCTTCTTACACCTACATG 95929  
QY 287 CTGACCGGTGAATGTGAGCATGTCTTCAAGCACTAACTTACCAATTAACCCAGAGGAA 346  
DB 95930 CCGACCGGTGAACGTGAGCATGTCTTCAAGCACTAACTTACCAATTAACCCAGAGGAA 95989  
QY 347 ----- 346  
DB 95990 AAGAACCATAGATCTTCACTGTAATGTAATAATGTGCTTGCACAGTACTACACTGACA 96049  
QY 347 -----TCGTGTACAGATCTTACATGAGCATGATGCTCTCT 378  
DB 96050 CAAAAAATGTCTGAAATATGACGGGTGAAGTGTACAGGTCTTACATGATGTGCTGCTC 96109  
QY 379 GGTGACGCATCTTCAACCGCCAGCGGAGCTGTGAGGAGCAGAGGAGACGCGAGT 438  
DB 96110 GGTGATGGCATATTCATATGCCAGCGGAGATGTGGAGGAACAAAGGAGACGCGGAGC 96169  
QY 439 TTCGAGTTTCCCTCCAAAGAACCTGAGGGATTTTTCAGCGCCATTTGTTCAGAGAGTACTCC 498  
DB 96170 TTCGAGTTTCCCTCCAAAGAACCTGAGAGACTTTCAGCACTGTGTTTTCAGGAGTACTCC 96229  
QY 499 CTGAAGCTGTGGGTGATGACGACGAGCATCCAAAGGACGAGGAAAGTGTGAGCATG--- 555  
DB 96230 CTGAAGCTATCAAGCATTTCTGAGCAAGCATGCAAGGCGCGGAGAGTTGTGACATGCG 96289  
QY 556 ----- 555  
DB 96290 GTAACCACTGAATTCCTTGCTTAATACCTTAACATTTCTTGAGAAACCAATTTGTCAG 96349  
QY 556 -----CAGGAACTTTTACATGAGGATGACGCTGAGCTC 587  
DB 96350 TAATCTGATGCAAGAACTAACCATAATTCAGAAATGTTTCATGAGATGACACTGACTC 96409  
QY 588 CATCTGCAAGTGTGGTTCGGGTGAGATCGGACGCTGTGCGCAGATCTCTCCCGAGAA 647  
DB 96410 GATCTGCAAGTGTGGGTTCGGGTGAGATCGGACGCTGTCACTGATCTCTCCCGAGAA 96469  
QY 648 CAGCTTCGCGCAGGCGTTCGATGCGCCCAACATCATCATGCTGCGGTTCATCGACCC 707  
DB 96470 CAGCTTTGCCAGGATTCGACGCTGCGCCAAATCATCATGCTGCGGTTCATCGATCC 96529  
QY 708 GCTGTGGGCACTCAAGAGTCTTCCAGTGGGTTCAGAGGCGCTCCTAGCGCAGAGCAT 767  
DB 96530 TCTGTGGGCTCTGAAGAAGTCTTTCAGCTCGATCAGAGGCTCTCTTCGACGAGCAT 96589  
QY 768 CAACTCTGTGACAGTTCACCTACAGCTGATTCGCGCGGAGAGGCGGAGATCGTGA 827  
DB 96590 GAACTGTGTTGATGACTTCACTACAGCTGATTCGCGCGGAGAGGCTGAGATCTTGA 96649  
QY 828 GGTTCGCGGCGCAGCGCAACAGGAG----- 853  
DB 96650 GGTTCGAGCAGCGGCAACAGAGAGGATGATCTTCTCTCTGCTCAAGAAATCAGT 96709  
QY 854 ----- 853  
DB 96710 AGAACTGAATGACATGGTAATGGTATGATCAGATCGGAAAGGTTTGTTCATTGATA 96769  
QY 854 -----AGTGAAGCAGCATCTCTGTACGGTTTCTGAGCTG 891  
DB 96770 TCGTTGATTTGTAATGGCGAGCAGATCAAGCAGCATATCTGTGCGGTTCATCGAGCTC 96829

QY 892 GCGAGGCGCGGAGGAC-----GGCGCGCTTCGGGAGGATAAAGCTCTCGGAC 945  
DB 96830 GGGAGGCGCGGCGGACGAGAGGGGGGCGCAGCTTCGGGGGACGAAAGCTCTCGCGAC 96899  
QY 946 GTGTGTCTCAACTTCTGATCGCGGGGCGGACACGAGCGGACGACGCTGTCTGTGGTTC 1005  
DB 96890 GTGTGTCTCAACTTCTGATCGCGGGGTCACACGAGCGGACGACGCTGTCTGTGGTTC 96949  
QY 1006 AGCACATGGCCATGTCCACCCCGAACCTGGCCGAGAAAGTGGCCGCGAGCTGTGCGCG 1065  
DB 96950 ACGTACATGGCGATGACACACCCCGCCGCTCGCCGACAAAGCTCCCGCGGAGCTGCGCGCG 97009  
QY 1066 TTCAGAGCGGACGCGCGCGGAGGAGCGCTGACGCTCGTCTGTGGCGGGCGCTGAC 1125  
DB 97010 TTCAGAGGATGAGCGCGCGCGGAGGAGCGTTCGCGCT-----CGCCGACCGCC 97057  
QY 1126 GCCACACGACAAAGGCGTTTCGCGCGCGCTGCGCAGTTTCGGGGGCTCTCTCACCTTACCGAC 1185  
DB 97058 GCCGCGAGGCGTCTGTCGCGCGCGCTGCGCAGTTTCGCGCTGCTGAGCTACGAC 97117  
QY 1186 AGCTTCGCAAGCTGTCTTACCTCACGCTGCTGTCACGAGACGCTCGGCTGTGACCC 1245  
DB 97118 GCGGTGGGAAAGCTGGTGTACTGTCACGCTGCTGTCAGGAGACGCTCGGCTCTTACCG 97177  
QY 1246 GCGGTTCCTCAGGACCCCAAGGGATCCTGAGGACGACGCTGTCGCGGACGCGGACGAG 1305  
DB 97178 GCGGTGCGCAGGACCCCAAGGGATCCTGAGGAGACGAGCTGCTCCCGACGCGACCAAG 97237  
QY 1306 GTGAGGCGCGCGGATGGTGAACCTGACCTGACCTGCTGCTGCTGAGGCGGATGAGTACAC 1365  
DB 97238 GTGCGCGCGCGGATGGTGAACCTGACCTGACCTGCTGCTGCTGAGGAGATGAGTACAC 97297  
QY 1366 TGGGCGCGCGGAGCGGAGCTTTCGCGCGGAGCGCTGTCGCGGAGCGCTGCTGCTG 1422  
DB 97298 TGGGCGCGCGGAGCGGAGCTTTCGCGCGGAGCGCTGCTGCGCGGAGCGCTGCTGCTG 97357  
QY 1423 TTCGCAACGCGTTCGCGCTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTT 1482  
DB 97358 TTCGCAACGCGTTCGCGCTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTT 97417  
QY 1483 GCGAAGGACTCGGCTTCACTGACATGAGATGCGCTGCGCTGCGCTGCGCTTCTTCCGCTTCTAC 1542  
DB 97418 GCGAAGGACTCGGCTTCACTGACATGAGATGCGCTGCGCTGCGCTTCTTCCGCTTCTAC 97477  
QY 1543 AGCTTCGCGCTGCTGAGGCGGACCCGCTGACGATGACGATGACGATGACGATGACGATG 1602  
DB 97478 ACCCTCGACCTCGTTCGAGGACCCCGCTCAAGTACCGGATGATGACCATCTCTTCCATG 97537  
QY 1603 GCGACGCGCTCAAGGTCGCGCTCTTAGGGCGCTCTGA 1641  
DB 97538 GCTCAGGCGCTCAAGGTCGCGCTCTTAGGGCGCTCTGA 97576

RESULT 5  
AK118803  
LOCUS Arabidopsis thaliana At1g9500 mRNA for unknown protein, complete  
DEFINITION cds, clone: RAF121-14-E14.  
ACCESSION AK118803  
VERSION AK118803.1 GI:26452619  
KEYWORDS FLI CDNA; CAP trapper.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1  
REFERENCE  
AUTHORS Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J.,  
Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P.,  
Kawai, J., Hayashizaki, Y. and Shinozaki, K.  
TITLE Arabidopsis thaliana full-length cDNA  
JOURNAL Published Only in Database (2002)  
REFERENCE 2 (bases 1 to 1695)

Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J.,  
 Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P.,  
 Kawai, J., Hayashizaki, Y. and Shinozaki, K.  
 Direct Submission  
 Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences  
 Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa  
 230-0045, Japan (E-mail: mseki@gsc.riken.go.jp,  
 URL: http://pfigweb.gsc.riken.go.jp, Tel: 81-45-503-9625,  
 Fax: 81-45-503-9586)  
 An Arabidopsis full-length cDNA library was constructed essentially  
 as reported previously (Seki et al. (1998) Plant J. 15:707-720;  
 Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI  
 and XhoI was ligated to modified Lambda FIC-1-E vector (Carninci et  
 al. (2001) Genomics 77:79-90) digested with BamHI and SalI.  
 This clone is in a modified pBluescript vector.  
 Please visit our web site (http://pfigweb.gsc.riken.go.jp/) for  
 further details.

FEATURES	Location/Qualifiers
source	1..1695 /organism="Arabidopsis thaliana" /mol_type="cDNA" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="1" /clone="RAF121-14-E14" /note="common name: thale cress"
gene	1..1695 /gene="Atlg69500/F10D13_15"
CDS	110..1546 /gene="Atlg69500/F10D13_15" /codon_start=1 /product="unknown protein" /protein_id="BAC43393.1" /db_xref="GI:26452620" /translation="MHDMLVEVLVNSRTVVVPMPTTYTVIADPINEVVLKTNESNY PRGETHYSYNEVLGLGFIENSDELWEKORKTASFASKNLRDFTSVFKYSLSKLF TILSASFKEQVDMQELLMLWFLDSICKVFGVEITGLAPELNEHFAKADTANII VTILRFIDPLQMKKKFIISEALLGKSIKVNDFTYSVTRRKAELNEAQISPTNNN NNNVKVHIDILRSFIEDSDPSKETEKLSDIVLNFVAGRTTATTITLWAIMLMH NNNKVASISELQELBEKSAEATNLSHQVDTEDSEFNKEKTEFAGLLNYSGLGKLH YLHAFITELRLVPAVQDPKGVLEDDMLPNGTKVKAGGMVTVVPYSGRMENYNGSD ALAFKPERVLKQGVFONASPFKTAFOAGPRICLGHKDSAYLQKMMALILCFYKPHL VNPHPVKYRNTTILSWAHGUKVTVRSRS"

ORIGIN	Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps	3;
QY	103	TCATGGATCTGTGCAGAGTGGAGCTGAGGAGCAAGCAAGCGCGGAGATCATGGCCA	162					
Db	8	TCATGGATCTTCTTCAACCGATGGGACAAAGAAACAAGAGTGGTCCAAAGACTTGGCCT	67					
QY	163	GTCAATCGGGCAACGGTGGAGCAGCTGAGGAACCTACCGGATGCACGACTGGCTTGTC	222					
Db	68	TTGGTTGGAGCAGCATCGACAGTTAACTAATTTGATCGATGATGATGGCTCGTT	127					
QY	223	GGGTACTGTACGGCAGCAGGACAGTGCAGTGCAGCATGCCGTTTCCTTCTACACTAC	282					
Db	128	GAGTATCTTTATAACTCAAGAACAGTAGTGGTTCCTAATGCCATCTCACTATTATACATAC	187					
QY	283	ATCGCTGACCGGTGAATGTCGAGCATGTCCTCAAGACTAACTTCACCAATTAACCCCAAG	342					
Db	188	ATAGCTGATCCCATCAACGTTGAATATGTCCTGAAAACAAATCTTCCAAATTAACCCCAAG	247					
QY	343	GGAAATCGTTCAGATCTTACATGGACGCTGCTCCTCGGTGACGGCATCTTCAACGCCGAC	402					
Db	248	GGAGAGACTTATCATTTTATATGGAAGTATCTTGGGAGATGGGATTTTCAATTTCTGAT	307					
QY	403	GGCAGCTGTGGAGGACGAGAGACGCGCAGTTTCGAGTTTCGCTTCCAGAACCTTG	462					
Db	308	GGAGAGCTTTGGAGGAAACAGAGAAAAACCGCAGTTTCGAATTTGCTTTCCAGAACTCTT	367					

```

Db      1439 AGATTCTATAGTTTCATTGTTGGTGCAGAAATCATCTCTGCAAGTATCGGATGATGACGATC 1498
QY      1594 CTCTCCATGGCGCAGCGCTCAAGTCCGCGTCTCTAG 1631
Db      1499 TTATCCATGGCTCATGGCTTGAAGTTACTGTATCCAG 1536

RESULT 6
LOCUS   AX453574
DEFINITION Sequence 3 from Patent WO0226789.
ACCESSION AX453574
VERSION   AX453574.1 GI:21712817
KEYWORDS Sorghum bicolor (sorghum)
SOURCE   Sorghum bicolor
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
AUTHORS  Albertsen,M.C., Fox,T., Huffman,G. and Trimmell,M.
TITLE    Nucleotide sequences mediating male fertility and method of using
JOURNAL  same
PIONEER HI-BRED INTERNATIONAL, INC. (US)
FEATURES
source   Patent: WO 0226789-A 3 04-APR-2002;
          Location/Qualifiers
          /organism="Sorghum bicolor"
          /mol_type="unassigned DNA"
          /db_xref="taxon:4558"

ORIGIN
Query Match 22.7%; Score 433.2; DB 6; Length 494;
Best Local Similarity 93.7%; Pred. No. 9.2e-45;
Matches 447; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY      259 ATGCGGTTACATCTCTACCTACCTACCTGCTGACCGGTGAATGTGAGCATGCTCTCAAG 318
Db      13 ATGCGGTTACATCTCTACCTACCTACCTGCTGACCGGTGAATGTGAGCATGCTCTCAAG 72

QY      319 ACTAACTTCAACAAATACCCAGGAATCGTGTACAGATCTCTACATGACGCTGCTCTC 378
Db      73 ACTAACTTCAACAAATACCCAGGGAGCTGTACAGATCTCTACATGATGCTCTCTC 132

QY      379 GTGACGGCATCTTCAACCGGACGGCGAGCTGTGGAGGAGCAGGAGGACGGCGAGT 438
Db      133 GTGACGGCATATTCAACGCTGACGGCGAGCTGTGGAGGAGCAGGAGGACGGCGAGT 192

QY      439 TTGAGTTTCGCTCCAAAGAACCTGAGGATTTTCAGGCCATTTGTTTCAGAGGACTCTCC 498
Db      193 TTGAGTTTCGCTCCAAAGAACCTGAGGATTTTCAGGCCATTTGTTTCAGAGGACTCTCC 252

QY      499 CTGAAGCTGTGGGTATCTAGTACGACGAGCATCCAGGACGCAAGTGTGACATGTCAG 558
Db      253 CTGAAGCTGTGGGTATCTAGTACGACGAGCATCCAGGACGCAAGTGTGACATGTCAG 312

QY      559 GAACCTTTACATGAGGATGACGCTGAGTCTCACTCTGCAAGTTGGGTTCGGGTTCGAGATC 618
Db      313 GAACCTTTACATGAGGATGACGCTGAGTCTCACTCTGCAAGTTGGGTTCGGGTTCGAGATC 372

QY      619 GGCAGGCTGTCCGAGATCTCCCGAGACAGCTTCGCGCAGCGTTCGATCCGCGCAC 678
Db      373 GGCAGGCTGTCCGAGATCTCCCGAGACAGCTTCGCGCAGCGTTCGATCCGCGCAC 432

QY      679 ATCATCATCAGCTGGGTTTCATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCAC 735
Db      433 ATCATCGTCACTGGGTTTCATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCAC 489

AK119612 1714 bp mRNA linear PLN 29-OCT-2003

```

```

Oryza sativa (japonica cultivar-group) cDNA clone.002-117-G01, full
insert sequence.
AK119612
GI:37989235
FLI cDNA; oligo capping.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.

1
The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team.,
Kikuchi,S., Saton,K., Nagata,T., Kawagashira,N., Doi,K.,
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ohtsuki,Y., Murakami,K.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN.,
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Yoshino,M. and Hayashizaki,Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
Science 301 (5631), 376-379 (2003)
2252273
12869764

2
ADACHI,J., AIZAWA,K., AKIMURA,T., ARAKAWA,T., CARNINCI,P., DOI,K.,
FUJIMURA,T., FUKUDA,S., HANAGAKI,T., HARA,A., HASHIZUME,W.,
HAYASHIDA,K., HAYASHIZAKI,Y., HAYATSU,N., HIRAMOTO,K., HIRAKA,T.,
HORI,F., HOTTA,I., IIDA,J., IIDA,Y., IKEDA,R., IMAMURA,K.,
IMOTANI,K., ISHIBIKI,J., ISHII,Y., ISHIKAWA,M., ITOH,M., KAGAWA,I.,
KANAGAWA,S., KATOH,H., KAWAGASHIRA,N., KAWAI,J., KAWAMATA,M.,
KIKUCHI,S., KISHIKAWA-HIROZANE,T., KISHIMOTO,N., KOBAYASHI,M.,
KODAMA,T., KOJIMA,K., KOJIMA,Y., KONDO,S., KONNO,H., KOUDE,M.,
KOYA,S., KURIHARA,C., KUROSAKI,T., KUSUMEGI,T., LI,C., LU,M.,
MASUDA,H., MATSUBARA,K., MATSUYAMA,T., MIURA,J., MIYAZAKI,A.,
MIZUNO,K., MURAKAMI,K., MURATA,M., NAGATA,T., NAKAHAMA,Y.,
NAKAMURA,M., NAMIKI,T., NARIKAWA,R., NIIKURA,J., NISHI,K.,
NOMURA,K., NUMASAKI,R., OHNEDA,E., OHNO,M., OHTSUKI,K., OKA,M.,
OOKA,H., OSATO,N., OTA,Y., OTOMO,Y., RYU,R., SAITOH,H., SAKAI,C.,
SAKAI,K., SAKAZUME,N., SANO,H., SASAKI,D., SATO,K., SATOH,K.,
SHIBATA,K., SHINAGAWA,A., SHIRAKI,T., SHISHIKI,T., SUGANO,Y.,
SUGANO,S., SUGIYAMA,A., SUZUKI,K., SUZUKI,Y., TAGAMI,M.,
TAGAMI-TAKEDA,Y., TAGAWA,A., TAKAHASHI,F., TAKAKU-AKIHARA,S.,
TANAKA,T., TOMARU,A., TOYA,T., TSUNODA,Y., UEDA,M., WAKI,K.,
XIE,Q., YAHAGI,W., YAMADA,H., YANAMOTO,M., YASUNISHI,A., YAZAKI,J.,
YOKOMIZO,S. and YOSHIMURA,A.
Collection, mapping, and annotation of 28K full-length cDNA clones
from japonica rice
Unpublished
3 (bases 1 to 1714)
Kikuchi,S.
Direct Submission
Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 32K full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cdna/
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K.,
Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T.,

```



## AUTHORS

The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ootomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kawaguchi, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Oka, K., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

Science 301 (5631), 376-379 (2003)

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

2 (bases 1 to 1780)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuka, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.

## TITLE

## JOURNAL

## COMMENT

Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>

NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., and Yamamoto, M.

FATS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,

Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.

## FEATURES

## source

1. 1780  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J03312LJ02"

## ORIGIN

Query Match 19.2%; Score 366.6; DB 8; Length 1780;  
Best Local Similarity 57.1%; Pred. No. 1.4e-36;  
Matches 821; Conservative 0; Mismatches 529; Indels 87; Gaps 5;

QY 141 GAAGGCCCCAGATCATGGCCAGTCATCCGCGCAACGGTGGAGCAGCTGAGGAACCTACCA 200  
DB 115 GAAGAAACCGAGGTATCCGCCCGTGGCGGCAACATGTTCCACAGTGTCTCAACTTCGG 174  
QY 201 CCGGATGCACGACTGGCTTGTCCGGTACCTGTCAACGCGCAGGACAGTGAACCTGCACAT 260  
DB 175 CCGCTTCTGGAGTACCAACAGGAGCTCTCCCGAAGTACCGCACCTTCGCGATGCTCAC 234  
QY 261 GCCCTTCACTTCTTACACCTTACATTCGGTGACCCGGTGAAATGTGAGCATGTCTCAAGAC 320  
DB 235 CCGGACCTGCAACTACGCTTACACCGTTCGAGCGCGCCACGCTGAGCACATCTCTCAAGAC 294  
QY 321 TAATCTTCACTTATCCCAAGGAAATCGTGATACAGATCTTACATGACCTGCTCTCTCGG 380  
DB 295 CAATCTTCGCAACTACGCGCAAGGCGCCGATGACCCAGCAGCTGTGGAGGACCTCTCTCGG 354  
QY 381 TGACGGCATCTTCAACCGCGCAGCGAGCTGTGGAGGAAGCAGAGGAGGAGGAGGAGTTT 440  
DB 355 CGACGGGATCTTCAACTGAGCGGCGCATGTGGCGGCGAGCGGCAAGGTTCGCGAGCT 414  
QY 441 CGAGTTCCCTTCCAAAGAACTGAGGGATTTTACGCCCATTTGTGTTCAGAGATCTCTCT 500  
DB 415 CGAGTTCTCCACCGTGTGTGCGGGACTACAGCAGCGCGGTGTTCGCGCACACCGCGC 474  
QY 501 GAAGCTGTGGGTATATCTG-----ACCGAGGATCCAGCGGCAAGTGTGGAGAT 554  
DB 475 GGAGTCCGCGGATCTCTGGAGCGTGTTCGCGGCGGCGAGGCGGCGGAGAGGTTGGACAT 534  
QY 555 GCAGAACTTTTACATGAGGATGACGCTGGACTTCCATCTGCAAGGTTCGGTTCGGGGTCA 614  
DB 535 GCAGGATCTGCTGATGCGGGCGAGCGTGGACTCTTTCTTTCAGGGTTGGTTTCGGGGTCAA 594  
QY 615 GATCGGACGCTGTCGCGCAGATCTCCCGGAGAAC---AGTTCGCGCAGGCGTTCGATGC 671  
DB 595 CATTGGCGTGTCTCTCCGGATCCAGCAAGGAGGGTGTGGTGTTCGCCAGGGCGTTCGACGA 654  
QY 672 CGCCAAATCATCATCAGCTGCGGTTCATCGACCCGCTGTGGCGCATCAGAGGTTCTT 731  
DB 655 CGCGAGGAGCAGGCTGTGTTCGCAATCTTTCGACCTGCTCTGGAAGGTCAAGAGTTTCT 714  
QY 732 CCAGCTCGGGTTCAGAGGCGCTCTCTAGCGCAGAGCATCAAGCTCTGGACGAGTTTCACCTA 791  
DB 715 CAACATCTCGTCGAGGCAACCATGAAGCAGTCCATCCGACCATCAACGACTTCGTGTA 774  
QY 792 CAGCGTGAATCCGCGGAGGAGGCGGAGATGCTGTCAGGTTTCGGGCCAGCGGCACACAGGA 851  
DB 775 CTCATCATTCACAGGA-----AGATCGAAACAGATGAGCAGAGAGCAACACGAAAT 825  
QY 852 GAAGATGAAGCAGCAGATCTCTGTCACGGTTTCATGAGCTGGCGGAGCGCGCGAGCGG 911  
DB 826 CGCCAAAGAGAGAGACATCTGTTCGAGGTTTCTGCTCGAGAGGGAG-----AGGA 876  
QY 912 CGGCGGCTTCGGGAGCAGATAGAGCCCTCCGGGAGCTGTGTGCTCAACTTCGTGATCGCGG 971  
DB 877 TCCCGGCTTCTCGAACCAAGTACATCAGGAGCATCATCACTCACTTCGTCATCGCGG 936

QY	972	GC	GGG	C	A	C	A	C	G	G	C	A	C	G	A	C	G	T	G	T	C	A	C	C	G	G	A	1031			
Db	937	CC	CG	C	A	C	A	C	G	C	G	G	A	C	G	T	G	T	C	T	C	T	C	A	C	C	A	G	A	996	
QY	1032	CG	T	G	G	C	G	A	A	G	T	G	C	G	C	G	A	G	T	G	T	G	C	G	G	T	T	C	G	A	1091
Db	997	CG	T	A	C	A	G	A	T	C	G	A	G	G	A	G	T	G	C	G	-----	-----	-----	-----	-----	-----	-----	-----	-----	1027	
QY	1092	GG	CG	T	C	A	C	G	T	C	T	G	C	G	C	G	C	T	G	A	C	C	G	C	A	C	A	A	G	C	1151
Db	1028	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1056	
QY	1152	CG	T	G	G	C	A	G	T	T	C	G	C	G	G	C	T	C	A	C	T	A	C	A	C	A	G	C	T	G	1211
Db	1057	CG	T	C	A	G	A	T	T	T	C	T	C	A	T	T	T	C	G	A	A	G	A	C	C	A	T	C	A	A	1116
QY	1212	CG	C	T	C	G	T	C	A	C	G	A	C	G	T	C	C	T	G	T	A	C	C	C	T	A	G	A	C	C	1271
Db	1117	CG	C	A	G	C	G	T	T	G	A	C	G	T	T	G	A	C	C	T	G	G	C	T	C	G	A	T	G	A	1176
QY	1272	C	T	T	G	A	G	A	C	G	A	G	T	G	C	C	G	A	C	G	G	A	G	T	G	A	G	C	G	A	1331
Db	1177	C	T	T	C	G	A	T	G	A	C	A	G	T	T	C	C	G	A	C	G	G	T	A	A	A	A	G	A	G	1236
QY	1332	CG	T	C	C	T	A	C	T	G	G	C	G	A	T	G	A	T	A	C	A	C	T	G	G	G	C	C	G	A	1391
Db	1237	C	A	A	C	G	T	A	C	C	C	A	T	G	G	C	A	G	A	T	G	A	T	T	C	T	G	T	G	G	1296
QY	1392	G	C	C	G	A	G	C	G	T	G	A	T	C	A	A	C	G	A	G	T	G	C	G	C	A	C	G	T	C	1451
Db	1297	G	C	C	G	A	G	C	G	T	G	C	T	C	G	A	T	G	T	T	C	G	C	G	A	G	C	C	C	T	1356
QY	1452	G	G	C	T	T	C	A	G	C	G	G	C	G	A	G	T	C	T	C	T	T	G	G	C	A	A	G	A	C	1511
Db	1357	G	A	C	T	T	C	A	G	C	G	G	C	C	A	C	A	T	C	T	G	T	G	G	A	G	A	G	T	T	1416
QY	1512	G	A	T	G	C	G	T	G	G	C	A	T	C	T	T	C	C	T	T	C	A	C	A	G	T	T	C	C	G	1568
Db	1417	G	A	T	C	G	T	T	C	G	G	T	G	C	T	T	C	T	P	A	C	T	T	C	A	G	A	T	T	C	1473

RESULT	9
LOCUS	AK101618
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone:J033052H03, full insert sequence.
ACCESSION	AK101618
VERSION	AK101618.1 GI:32986827
KEYWORDS	FLI CDNA; CAP trapper.
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Erbartoideae; Oryzaceae; Oryza.
REFERENCE	1
AUTHORS	The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto.N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka.H., Hotta,I., Koijima.K., Namiki.T., Ohneda,E., Yahagi.W., Suzuki.K., Li.C., Ohtsuki.K., Shishiki.T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group:, Otomo.Y., Murakami.K., Iida,Y., Sugano.S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Narusaka,T., Kodama.T., Masuda,H., Kobayashi.M., Xie,Q., Lu,M., Narikawa,R., Sugiyaama,A., Mizuno,K., Yokomizo,S., Niikura.J., Ikeda.R., Ishibiki,J., Kawanata.M., Yoshimura.A., Miura,J., Kusumegi,T., Oka.M., Ryu.R., Ueda.M., Matsubara.K., RIKEN. Kawai.J., Carninci.P., Adachi,J., Aizawa.K., Arakawa,T., Fukuda.S., Hara.A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii.Y., Itoh,M., Kagawa.I., Kondo,S., Konno.H., Miyazaki,A., Osato,N., Ota.Y., Saito.R., Sasaki,D., SATO,K., Shibata.K., Shinagawa,A., Shiraki.T., Yoshino.M. and Hayashizaki,Y.



141 GAAAGCCCGAGATCATGGCCAGTATCGGCGCAACGGTGGAGCAGCTGAGAACTACCA 200  
142 |||||  
143 |||||  
144 |||||  
145 |||||  
146 |||||  
147 |||||  
148 |||||  
149 |||||  
150 |||||  
151 |||||  
152 |||||  
153 |||||  
154 |||||  
155 |||||  
156 |||||  
157 |||||  
158 |||||  
159 |||||  
160 |||||  
161 |||||  
162 |||||  
163 |||||  
164 |||||  
165 |||||  
166 |||||  
167 |||||  
168 |||||  
169 |||||  
170 |||||  
171 |||||  
172 |||||  
173 |||||  
174 |||||  
175 |||||  
176 |||||  
177 |||||  
178 |||||  
179 |||||  
180 |||||  
181 |||||  
182 |||||  
183 |||||  
184 |||||  
185 |||||  
186 |||||  
187 |||||  
188 |||||  
189 |||||  
190 |||||  
191 |||||  
192 |||||  
193 |||||  
194 |||||  
195 |||||  
196 |||||  
197 |||||  
198 |||||  
199 |||||  
200 |||||  
201 CCGGATGACGAGCTGGCTTGTGCGGTACCTGTCACGGCAGAGGACAGTGCCTCGAT 260  
202 |||||  
203 |||||  
204 |||||  
205 |||||  
206 |||||  
207 |||||  
208 |||||  
209 |||||  
210 |||||  
211 |||||  
212 |||||  
213 GAGAAACGCGAGGTATCGCCCGTGGCGGCGACCATGTTCCACAGCTGCTCAACTTCGG 272  
214 |||||  
215 |||||  
216 |||||  
217 |||||  
218 |||||  
219 |||||  
220 |||||  
221 |||||  
222 |||||  
223 |||||  
224 |||||  
225 |||||  
226 |||||  
227 |||||  
228 |||||  
229 |||||  
230 |||||  
231 |||||  
232 |||||  
233 |||||  
234 |||||  
235 |||||  
236 |||||  
237 |||||  
238 |||||  
239 |||||  
240 |||||  
241 |||||  
242 |||||  
243 |||||  
244 |||||  
245 |||||  
246 |||||  
247 |||||  
248 |||||  
249 |||||  
250 |||||  
251 |||||  
252 |||||  
253 |||||  
254 |||||  
255 |||||  
256 |||||  
257 |||||  
258 |||||  
259 |||||  
260 |||||  
261 GCCGTTCATCTTCACTACATCATGCTGACCCGGTGAATGTCCAGCATGTCCTCAAGAC 320  
262 |||||  
263 |||||  
264 |||||  
265 |||||  
266 |||||  
267 |||||  
268 |||||  
269 |||||  
270 |||||  
271 |||||  
272 |||||  
273 CCGCTTGTGAGTACCAACACGAGAGCTCTCCCGAAGTACCGACCTTCCTCCGATGCTCAC 332  
274 |||||  
275 |||||  
276 |||||  
277 |||||  
278 |||||  
279 |||||  
280 |||||  
281 |||||  
282 |||||  
283 |||||  
284 |||||  
285 |||||  
286 |||||  
287 |||||  
288 |||||  
289 |||||  
290 |||||  
291 |||||  
292 |||||  
293 |||||  
294 |||||  
295 |||||  
296 |||||  
297 |||||  
298 |||||  
299 |||||  
300 |||||  
301 |||||  
302 |||||  
303 |||||  
304 |||||  
305 |||||  
306 |||||  
307 |||||  
308 |||||  
309 |||||  
310 |||||  
311 |||||  
312 |||||  
313 |||||  
314 |||||  
315 |||||  
316 |||||  
317 |||||  
318 |||||  
319 |||||  
320 |||||  
321 TAACCTTCCAAATTAACCCAGAGGAATCGTGTAAGATCCCTACATGAGAGTGCCTCGG 380  
322 |||||  
323 |||||  
324 |||||  
325 |||||  
326 |||||  
327 |||||  
328 |||||  
329 |||||  
330 |||||  
331 |||||  
332 |||||  
333 |||||  
334 |||||  
335 |||||  
336 |||||  
337 |||||  
338 |||||  
339 |||||  
340 |||||  
341 |||||  
342 |||||  
343 |||||  
344 |||||  
345 |||||  
346 |||||  
347 |||||  
348 |||||  
349 |||||  
350 |||||  
351 |||||  
352 |||||  
353 |||||  
354 |||||  
355 |||||  
356 |||||  
357 |||||  
358 |||||  
359 |||||  
360 |||||  
361 |||||  
362 |||||  
363 |||||  
364 |||||  
365 |||||  
366 |||||  
367 |||||  
368 |||||  
369 |||||  
370 |||||  
371 |||||  
372 |||||  
373 |||||  
374 |||||  
375 |||||  
376 |||||  
377 |||||  
378 |||||  
379 |||||  
380 |||||  
381 TGACGGCATCTTCAACGCCGACGGAGCTGTGAGGAGAGCAGAGGAGAGCGGAGTTT 440  
382 |||||  
383 |||||  
384 |||||  
385 |||||  
386 |||||  
387 |||||  
388 |||||  
389 |||||  
390 |||||  
391 |||||  
392 |||||  
393 |||||  
394 |||||  
395 |||||  
396 |||||  
397 |||||  
398 |||||  
399 |||||  
400 |||||  
401 |||||  
402 |||||  
403 |||||  
404 |||||  
405 |||||  
406 |||||  
407 |||||  
408 |||||  
409 |||||  
410 |||||  
411 |||||  
412 |||||  
413 |||||  
414 |||||  
415 |||||  
416 |||||  
417 |||||  
418 |||||  
419 |||||  
420 |||||  
421 |||||  
422 |||||  
423 |||||  
424 |||||  
425 |||||  
426 |||||  
427 |||||  
428 |||||  
429 |||||  
430 |||||  
431 |||||  
432 |||||  
433 |||||  
434 |||||  
435 |||||  
436 |||||  
437 |||||  
438 |||||  
439 |||||  
440 |||||  
441 CGAGTTTCGCTCCAGAACTTGGAGGATTTTCAGCGCCATTGTTTTCAGAGAGTATCCCT 500  
442 |||||  
443 |||||  
444 |||||  
445 |||||  
446 |||||  
447 |||||  
448 |||||  
449 |||||  
450 |||||  
451 |||||  
452 |||||  
453 |||||  
454 |||||  
455 |||||  
456 |||||  
457 |||||  
458 |||||  
459 |||||  
460 |||||  
461 |||||  
462 |||||  
463 |||||  
464 |||||  
465 |||||  
466 |||||  
467 |||||  
468 |||||  
469 |||||  
470 |||||  
471 |||||  
472 |||||  
473 |||||  
474 |||||  
475 |||||  
476 |||||  
477 |||||  
478 |||||  
479 |||||  
480 |||||  
481 |||||  
482 |||||  
483 |||||  
484 |||||  
485 |||||  
486 |||||  
487 |||||  
488 |||||  
489 |||||  
490 |||||  
491 |||||  
492 |||||  
493 |||||  
494 |||||  
495 |||||  
496 |||||  
497 |||||  
498 |||||  
499 |||||  
500 |||||  
501 GAACTGTGCGGTATATCTG-----AGCCAGGCAATCCAAAGCAGCAAGTGTGGACAT 554  
502 |||||  
503 |||||  
504 |||||  
505 |||||  
506 |||||  
507 |||||  
508 |||||  
509 |||||  
510 |||||  
511 |||||  
512 |||||  
513 |||||  
514 |||||  
515 |||||  
516 |||||  
517 |||||  
518 |||||  
519 |||||  
520 |||||  
521 |||||  
522 |||||  
523 |||||  
524 |||||  
525 |||||  
526 |||||  
527 |||||  
528 |||||  
529 |||||  
530 |||||  
531 |||||  
532 |||||  
533 |||||  
534 |||||  
535 |||||  
536 |||||  
537 |||||  
538 |||||  
539 |||||  
540 |||||  
541 |||||  
542 |||||  
543 |||||  
544 |||||  
545 |||||  
546 |||||  
547 |||||  
548 |||||  
549 |||||  
550 |||||  
551 |||||  
552 |||||  
553 |||||  
554 |||||  
555 GCAGGAATTTTACATGAGATGACGCTGATCTCCATCTGCAAGGTTGGTTTCGGGGTCTGA 614  
556 |||||  
557 |||||  
558 |||||  
559 |||||  
560 |||||  
561 |||||  
562 |||||  
563 |||||  
564 |||||  
565 |||||  
566 |||||  
567 |||||  
568 |||||  
569 |||||  
570 |||||  
571 |||||  
572 |||||  
573 |||||  
574 |||||  
575 |||||  
576 |||||  
577 |||||  
578 |||||  
579 |||||  
580 |||||  
581 |||||  
582 |||||  
583 |||||  
584 |||||  
585 |||||  
586 |||||  
587 |||||  
588 |||||  
589 |||||  
590 |||||  
591 |||||  
592 |||||  
593 |||||  
594 |||||  
595 |||||  
596 |||||  
597 |||||  
598 |||||  
599 |||||  
600 |||||  
601 |||||  
602 |||||  
603 |||||  
604 |||||  
605 |||||  
606 |||||  
607 |||||  
608 |||||  
609 |||||  
610 |||||  
611 |||||  
612 |||||  
613 |||||  
614 |||||  
615 |||||  
616 |||||  
617 |||||  
618 |||||  
619 |||||  
620 |||||  
621 |||||  
622 |||||  
623 |||||  
624 |||||  
625 |||||  
626 |||||  
627 |||||  
628 |||||  
629 |||||  
630 |||||  
631 |||||  
632 |||||  
633 |||||  
634 |||||  
635 |||||  
636 |||||  
637 |||||  
638 |||||  
639 |||||  
640 |||||  
641 |||||  
642 |||||  
643 |||||  
644 |||||  
645 |||||  
646 |||||  
647 |||||  
648 |||||  
649 |||||  
650 |||||  
651 |||||  
652 |||||  
653 |||||  
654 |||||  
655 |||||  
656 |||||  
657 |||||  
658 |||||  
659 |||||  
660 |||||  
661 |||||  
662 |||||  
663 |||||  
664 |||||  
665 |||||  
666 |||||  
667 |||||  
668 |||||  
669 |||||  
670 |||||  
671 |||||  
672 |||||  
673 |||||  
674 |||||  
675 |||||  
676 |||||  
677 |||||  
678 |||||  
679 |||||  
680 |||||  
681 |||||  
682 |||||  
683 |||||  
684 |||||  
685 |||||  
686 |||||  
687 |||||  
688 |||||  
689 |||||  
690 |||||  
691 |||||  
692 |||||  
693 |||||  
694 |||||  
695 |||||  
696 |||||  
697 |||||  
698 |||||  
699 |||||  
700 |||||  
701 |||||  
702 |||||  
703 |||||  
704 |||||  
705 |||||  
706 |||||  
707 |||||  
708 |||||  
709 |||||  
710 |||||  
711 |||||  
712 |||||  
713 |||||  
714 |||||  
715 |||||  
716 |||||  
717 |||||  
718 |||||  
719 |||||  
720 |||||  
721 |||||  
722 |||||  
723 |||||  
724 |||||  
725 |||||  
726 |||||  
727 |||||  
728 |||||  
729 |||||  
730 |||||  
731 |||||  
732 |||||  
733 |||||  
734 |||||  
735 |||||  
736 |||||  
737 |||||  
738 |||||  
739 |||||  
740 |||||  
741 |||||  
742 |||||  
743 |||||  
744 |||||  
745 |||||  
746 |||||  
747 |||||  
748 |||||  
749 |||||  
750 |||||  
751 |||||  
752 |||||  
753 |||||  
754 |||||  
755 |||||  
756 |||||  
757 |||||  
758 |||||  
759 |||||  
760 |||||  
761 |||||  
762 |||||  
763 |||||  
764 |||||  
765 |||||  
766 |||||  
767 |||||  
768 |||||  
769 |||||  
770 |||||  
771 |||||  
772 |||||  
773 |||||  
774 |||||  
775 |||||  
776 |||||  
777 |||||  
778 |||||  
779 |||||  
780 |||||  
781 |||||  
782 |||||  
783 |||||  
784 |||||  
785 |||||  
786 |||||  
787 |||||  
788 |||||  
789 |||||  
790 |||||  
791 |||||  
792 |||||  
793 |||||  
794 |||||  
795 |||||  
796 |||||  
797 |||||  
798 |||||  
799 |||||  
800 |||||  
801 |||||  
802 |||||  
803 |||||  
804 |||||  
805 |||||  
806 |||||  
807 |||||  
808 |||||  
809 |||||  
810 |||||  
811 |||||  
812 |||||  
813 |||||  
814 |||||  
815 |||||  
816 |||||  
817 |||||  
818 |||||  
819 |||||  
820 |||||  
821 |||||  
822 |||||  
823 |||||  
824 |||||  
825 |||||  
826 |||||  
827 |||||  
828 |||||  
829 |||||  
830 |||||  
831 |||||  
832 |||||  
833 |||||  
834 |||||  
835 |||||  
836 |||||  
837 |||||  
838 |||||  
839 |||||  
840 |||||  
841 |||||  
842 |||||  
843 |||||  
844 |||||  
845 |||||  
846 |||||  
847 |||||  
848 |||||  
849 |||||  
850 |||||  
851 |||||  
852 |||||  
853 |||||  
854 |||||  
855 |||||  
856 |||||  
857 |||||  
858 |||||  
859 |||||  
860 |||||  
861 |||||  
862 |||||  
863 |||||  
864 |||||  
865 |||||  
866 |||||  
867 |||||  
868 |||||  
869 |||||  
870 |||||  
871 |||||  
872 |||||  
873 |||||  
874 |||||  
875 |||||  
876 |||||  
877 |||||  
878 |||||  
879 |||||  
880 |||||  
881 |||||  
882 |||||  
883 |||||  
884 |||||  
885 |||||  
886 |||||  
887 |||||  
888 |||||  
889 |||||  
890 |||||  
891 |||||  
892 |||||  
893 |||||  
894 |||||  
895 |||||  
896 |||||  
897 |||||  
898 |||||  
899 |||||  
900 |||||  
901 |||||  
902 |||||  
903 |||||  
904 |||||  
905 |||||  
906 |||||  
907 |||||  
908 |||||  
909 |||||  
910 |||||  
911 |||||  
912 |||||  
913 |||||  
914 |||||  
915 |||||  
916 |||||  
917 |||||  
918 |||||  
919 |||||  
920 |||||  
921 |||||  
922 |||||  
923 |||||  
924 |||||  
925 |||||  
926 |||||  
927 |||||  
928 |||||  
929 |||||  
930 |||||  
931 |||||  
932 |||||  
933 |||||  
934 |||||  
935 |||||  
936 |||||  
937 |||||  
938 |||||  
939 |||||  
940 |||||  
941 |||||  
942 |||||  
943 |||||  
944 |||||  
945 |||||  
946 |||||  
947 |||||  
948 |||||  
949 |||||  
950 |||||  
951 |||||  
952 |||||  
953 |||||  
954 |||||  
955 |||||  
956 |||||  
957 |||||  
958 |||||  
959 |||||  
960 |||||  
961 |||||  
962 |||||  
963 |||||  
964 |||||  
965 |||||  
966 |||||  
967 |||||  
968 |||||  
969 |||||  
970 |||||  
971 |||||  
972 |||||  
973 |||||  
974 |||||  
975 |||||  
976 |||||  
977 |||||  
978 |||||  
979 |||||  
980 |||||  
981 |||||  
982 |||||  
983 |||||  
984 |||||  
985 |||||  
986 |||||  
987 |||||  
988 |||||  
989 |||||  
990 |||||  
991 |||||  
992 |||||  
993 |||||  
994 |||||  
995 |||||  
996 |||||  
997 |||||  
998 |||||  
999 |||||  
1000 |||||

Qy 1212 CGCTCGTCCACCGAGACGCTCCGCTGTATACCCCGCGCTCTCAGAACCCCAAGGGGAT 1271  
Db |||||  
1215 CGCAGCTTTCACGAGACGCTCCGGTTGTATACCTGGCGTTTCCCTCGATGTCAAATCTG 1274  
Qy 1272 CCTGGAGACGAGCTGCTGCCGACGAGAGTGTAGGCGCGGGGATGGTGACGTA 1331  
Db |||||  
1275 CTTCTCGATGACACAGTTGCCGACGGGACGCGGTGAAGAAAGAGACATGGTGAAC 1334  
Qy 1332 CGTGCCCTACTCGATGGGGGGATGGAGTACAACTGGGGGGCCCGACGCGGAGCTTCCG 1391  
Db |||||  
1335 CCAACCGTACCCCTACCGATGGCAGGATGAAGTTCTCTGTGGGCGACGACGCCGAGGAT 1394  
Qy 1392 GCGGACGCGTGTATCAACGAGGATGGCGGTTCGCGCAACGCGTTCGCGTTCAAGTTCA 1451  
Db |||||  
1395 GCGGACGCGTGGCTCGATGACAGTGTATGTTCTGTCGCGGAGAGCCCTTCAAGTTCA 1454  
Qy 1452 GCGCTTCCAGCGCGGGCGGAGGATCTCGCTGGGCAAGGACTCGGCGTACCTGCAGATGAA 1511  
Db |||||  
1455 GCGCTTCCAGCGGGGGCGACGAATCTGCTGGGGAAGAGTTCGCTACAGGCGAGATGAA 1514  
Qy 1512 GATGCGCTGGCCATCTCTTCCGCTTCTACAGCTTCGCGTGTGGAGGGGCACCC 1568  
Db |||||  
1515 GATGCTTTCGCGTCTCTCTTCTACTTCTCAGATTCGAGATGTGGGATGACGAGCG 1571

## RESULT 10

AK120901

LOCUS

DEFINITION

Oryza sativa (japonica cultivar-group) 1712 bp mRNA linear PLN 29-OCT-2003

insert sequence.

AK120901

ACCESSION

AK120901.1 GI:37990524

VERSION

F01\_CDNA; CAP trapper.

KEYWORDS

Oryza sativa (japonica cultivar-group)

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1

The Rice Full-length cDNA Consortium, National Institute of

Agrobiological Sciences Rice Full-length cDNA Project Team,

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,

Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,

Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,

Otsuki, K., Shishiki, T., Foundation of Advancement of International

Science Genome Sequencing &amp; Analysis Group: Otsu, Y., Murakami, K.,

Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,

Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,

Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,

Ireda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,

Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,

Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,

Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Iehii, Y., Itoh, M.,

Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,

Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,

Yoshino, K. and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from

japonica rice

Science 301 (5631), 376-379 (2003)

JOURNAL

MEDLINE

22752273

PUBMED

12869764

REFERENCE

2

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,

Fujimura, T., Fukuda, S., Hangaki, T., Hara, A., Hashizume, W.,

Hayashida, K., Hayashizaki, Y., Hayateu, N., Hiramoto, K., Hiraoka, T.,

Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Inamura, K.,

Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,

Kawagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,

Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,

Koda, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,

Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,

Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakahama, Y.,



TITLE	Collection, mapping, and annotation of 28K full-length cDNA clones from japonica rice
JOURNAL REFERENCE	3 (bases 1 to 1712)
AUTHORS	Kikuchi S.
TITLE	Direct Submission
JOURNAL	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@ias.affrc.go.jp). Tel: 81-29-838-7007, Fax: 81-29-838-7007
COMMENT	This clone is one of the 32K full-length cDNA clones from japonica rice.

Db	234	CCCGACCTGCAACTACGTTCTACACCGTTCGAGCGCGCCAAACGTGCAGACACATCTCTCAAGAC	293
Qy	321	TAACTTTACCAATTAACCCCAAGGAATTCGTTATCAGATCTCTACATGACGCTGCTCTCCG	380
Db	294	CAACTTCGCCAACTACGGCAAGGCGCGATGACCCACGACGTCGTGAGGACCTCTCCG	353
Qy	381	TGAGCGCATCTTCAACGCCGACGGCGNCTGTGAGGAGACAGAGGAGGACGCGGAGTTT	440
Db	354	CGACGGGATCTTCAACGTGACGGCGCATGTGGCGGCACGCGCAAGGTGCCAGCCT	413
Qy	441	CGAGTTTCGCTTCAAGAACCTGAGGAGTTTTCAGCGCCATTGTGTTTCAGAGTACTCCCT	500
Db	414	CGAGTTCTCCACCGTGTCTGCGGACATACAGCAGCGCGTGTTCGCGCACCCGCGCG	473
Qy	501	GAAGCTGTCCGGTATACTG-----AGCCAGGCATCCAAGGCAAGGCAAGTTTGTGACAT	554
Db	474	GGAGCTGCGCGGCATCTCTGGAGCGTGGTCGCGCGCGGAAGGGGCGGAGAGGTTGGACAT	533
Qy	555	GCAGGAACCTTATCATGAGGATGACGTGGGACTCCATCTGCAAGTTTGGGTTTCGGGGTCGA	614
Db	534	GCAGGATCTGCTGATGGGCGACGCTGGACTCTTCTTCAGGGTTGGTTTCGGGGTCAA	593
Qy	615	GATCGGCAOCGTGTGCGCAGATCTCCCGGAGAAC--AGTTCGCGCAGCGGCTTCGATGC	671
Db	594	CCTTGGCGTCTCTCCGGATCCAGCAAGGAGGGGTGGTGTGTTGCCAGGCGTTTCGACGA	653
Qy	672	CGCCAAACATCATCACGCTCGGGTTCATCGACCCGCTGTCGGCGCATCAAGAGTTCTT	731
Db	654	CGCGAGGACGAGTGTGTTCGATCTCTTCGACCTGCTCTGGAAGTCAAGAGTTTCT	713
Qy	732	CCAAGTCGGGTACAGAGCCCTCTTAGCGCAGAGCATCAAGCTCGTCGACAGTTCACTTA	791
Db	714	CAACATCTCGTCGGAGGCAACCATGAAGCAGTCGATCCGCACCATCAACGACTTCGTGTA	773
Qy	792	CAGGTCATCCGCGGAGGAGGCCGAGATCGTTCGAGTCCGGGCCAGCGGCAACAGGA	851
Db	774	CTCCATCATCGACAGGA-----AGATCGAACAGATGAGCAGAGAGCAACACGAATT	824
Qy	852	GAAGATGAAGCAGACATCTGTCAAGTTTCATCGAGCTGGCGAGGCGCCGACGACGCG	911
Db	825	CGCCAGAAGAGGACATATCTGTCGAGTTTCTGCTCGAGGGAG-----AAGGA	875
Qy	912	CGCGGGTTCGGGACGATGAAGAGCTCCCGGAGTGGTGCTCAACTGTTGATCGCGCG	971
Db	876	TCCCGGCTGTTCGACAAAGTATACACAGGACATCATACTCAACTTCGTCTATCGCCGG	935
Qy	972	GCGGGACACACGCGCAGCAGCTGTCTGTGTTTCAAGCAGATGCCCATGTCCCAACCCGGA	1031
Db	936	CCGCGACACGACGCGGGGACGCTGTCTGTGTTTCTCTACGCCGTCTGCAAGAACCAAG	995
Qy	1032	CGTGGCCGAGAAATGCGCGCAGCTGTGCGGTTTCGAGGCGAGCGCGCGCGCAGGA	1091
Db	996	CGTACAGGACAAGATCGCAGGGAAATGCGC-----	1026
Qy	1092	GGGCGTCACGCTCGTGTCTGTCGCGCGCGCTGAGCCGACGACGAAGGCGTTTCGCGCCCG	1151
Db	1027	-----GAGCCACACCGCGACCGCGACGTCGCG	1055
Qy	1152	CGTGGCGAGTTTCGCGGGCTCTCTCACTTACGACAGCTTCGGCAAGCTGGTCTACCTCCA	1211
Db	1056	CGTCCAGGATTTCTCTTCAATTTCTGTGAGAAGACGCCATCAAAGAATGCAGTACCTACA	1115
Qy	1212	CGCTGCGTCAACCGAGACGCTCCGCTGTACCCGCGCTCCCTCAGGACCCCAAGGGAT	1271
Db	1116	CGCAGCGTTACGGAGACGCTCCGTTGTATCCCTGCGCTTCCCTCGATGTCAATACTG	1175
Qy	1272	CCTGGAGGACGACGTGCTGCGGACGGGACGAAGAGTGAAGGCGCGGGATGGTGACGTA	1331
Db	1176	CTTCTCGGATGACACTGTTCGGACGCGGACGCGGTGGAAGAGGACATGTTGACTA	1235
Qy	1332	CGTGCCCTACTCGATGGGCGGATGAGTACAATGGGGCCCCGACGCGGCGAGCTTCG	1391
Db	1236	CCAAACCTACCCATGGGACGATGAAGTTCTGTGGGGCGACGACGCCGAGAGTTCAA	1295

FEATURES SOURCE

## ORIGIN

	Query Match	19.1%;	Score 363.4;	DB 8;	Length 1712;
	Best Local Similarity	57.0%;	Pred. No. 3.4e-36;		
	Matches 819;	Conservative 0;	Mismatches 531;	Indels 87;	Gaps 5;
Qy	141	GAAGGCCGAGATCATGGCCAGTCTATGGCGCAACGGTGGAGCAGCTGAGGAATCA	200		
Db	114	GAGAAACGCGAGGTATCCGCCGTGGCGGCACCATGTTCACAGCTGCTCACTTCGG	173		
Qy	201	CCGATGCACGACTGGCTTGTCCGGTACCTGTACGGCACAGGACAGTGCACAT	260		
Db	174	CCGCTTCTGAGTAAACACACGAGGCTCTCCCGCAAGTACCGCACCTTCGCGATGCTCAC	233		
Qy	261	GCCTTCACTTCTTACACCTATCATCGCTGACCCCGGTGAATGTCAGCATGTCCTCAAGAC	320		



QY 497 CCCTGAAGCTCTCGGTTATATGAGCCAGGCGATCAAGGAGGAAAGTTGTGACATGC 556  
Db 589 CCGCAAGCTCGCGGCGTCTCCAGCCAGCGCGCTCAAAACCAATCCATGACATTC 648  
QY 557 AGGAACCTTATCATGAGGATGAGCTGGACTCCATCTGCAAGTTGGTTGGGTCGAGA 616  
Db 649 AGGTTTCTGATGAGAGCAACGATGGACTCCATCTTCCACCATCGCTGGCGCAAGACC 708  
QY 617 TCAGCAGCTCTCGCAGATCTCCCGAGAGACAGCTTCGCGAGCGCTTCGATCGCGCA 676  
Db 709 TCAACACGCTCGAGCTCGCGAGGCGCGCTTCGCGCGCGCTTCGACAGCCCA 768  
QY 677 ACATCATCATCAGCTGGGTTATCGACCGCTGTGCGCGATCAAGAGTTCTTCACG 736  
Db 769 GCGAGTTTACCATGCTCGCTACCTCAACCGCTTCTGGAAGCTTCGAGGCTCTCAACG 828  
QY 737 TCAGGTCAGAGGCGCTCTAGCGAGAGCATCAAGCTCTGAGAGTTCACCTACAGCG 796  
Db 829 TCAGGCGGAGCGATGCTCAAGAGAGGATCAAGTCTGTCGAGGCTTCGTACAGC 889  
QY 797 TGATCCGCGGAGGAGCGAGATCGTTCGAGGTCCGCGCGAGCGGCAACAGGAGAGA 856  
Db 889 TCATCCGTCAGAGCTCGAGAGCTCTCAACACCAAGGCAACGACACTGATTCGAG-- 946  
QY 857 TGAAGCAGCATCTCTGTCAGGTTTCATCGAGCTGGGGAGGCGCGGAGCGAGCGGCG 916  
Db 947 ----GCAGGATATCTGACAGATTTCAT-----CCAGGCAACGACTAGCGATTCTG 993  
QY 917 GCTTCGGGCGAGTAAGAGCTCCGCGGACGTGTGCTCAACTCGTGTATCGCGGCGGG 976  
Db 994 GGACGGTTGATTCAAGTACTTGAGAGACATATTTGAACATTTCTATAGCGCGAAG 1053  
QY 977 ACAGCGGCGACAGCGTGTGTTGTCAGCATGAGCATGAGGCAATCCCAACCGGACGTG 1036  
Db 1054 ACACACAGCGGCTGCTGTTGTTGTTCTGTACATGATGTGCAACACCGGAGTAC 1113  
QY 1037 CCAGAGAGCTCGCGGAGCTGTGCGGTTCGAGGCGGAGCGCGCGGAGGAGCGG 1096  
Db 1114 AGGAAGATCTGCCACA-----1132  
QY 1097 TCAGCTCTGCTCTCGCGGCGGCTGAGCGCGGACCAAGCGCTTCGCGCGCGGTGG 1156  
Db 1133 -----AGCATGAGGCGCCACACGCGCGGAGCGCTTCATCG 1173  
QY 1157 CGAGTTTCGGGCGCTCTCACCTACAGACCTTCGCGCAAGCTGTCTACCTCCAGCCT 1216  
Db 1174 ACAGATTCTCGCAGAGCTGACCGAGGCACTGAAACAGATGCACTATCTGACGCTG 1233  
QY 1217 GCGTACCGAGAGCTCGGCTGTACCCCGCGCTCCCTCAGGACCCCAAGGGATCTGG 1276  
Db 1234 CACTGACGAGAGCTCAGGCTATPACCTCAGTTCCACTGGATTAACAAGCAGTGTCT 1293  
QY 1277 AGGACGCTGCTCGCGGACGGACGAAGTGTAGGGCGCGGAGTGGTACGCTACGTCG 1336  
Db 1294 CAGACGATGATTGCGCAACGCGATTCAACGCTCAGCAGGGGACATCGTTCTTATCTC 1353  
QY 1337 CTTACTTCATGGGCGGATGAGTACAACTGGGCGCCCGACGCGGAGCTTCGCGCGCG 1396  
Db 1354 CTTACGCGATGGGCGGATGAGAGCTTGTGGGCGAAGACGCTGAATCTCTCCGCGCTG 1413  
QY 1397 AGCGGTGATCAAGAGGATGGCGCTTCGCAACGCTCGCGTTCAAGTTACGCGCT 1456  
Db 1414 ACGTTGGCTCGATGAAACGCGCTCTTTTTCAGCAGGAGACCGCTTCAATTTACAGCTT 1473  
QY 1457 TCAGGCGGCGGAGGATCTGCTGGCAAGGACTCGCGGTACCTGCGAGATGAAGATGG 1516  
Db 1474 TCAGGCGGCGGCAAGATCTGCTCGGAGAGATTTCGGTACAGCAGATGAAGATCT 1533  
QY 1517 CGTGGCATCTCTTCGCTTCTACAGCTTCGCGTTCGCTGAGGAGGCGACCGCG---GTGC 1573  
Db 1534 TCAGGCGGCTGCTCTCGCTTCTTCTGCTGCTCAAGCTCGCGGACGAGAGGATCATCA 1593

QY 1574 AGTACCGCATGATGACCATCTCTCCATGCGGACGCGCTCAA 1616  
Db 1594 GCTACCGGACCATGATTACACTCTCCGTCGATCAGGCTCTCCA 1636  
RESULT 12  
AK099925  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) cDNA clone:J013116M22, full insert sequence.  
ACCESSION  
AK099925.1 GI:32985134  
VERSION  
AK099925  
KEYWORDS  
F11\_CDNA; CAP trapper.  
SOURCE  
Oryza sativa (japonica cultivar-group)  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE  
1  
The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yanagi, M., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Ohtsuki, K., Tada, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Nariyawa, R., Sugiyama, A., Mizuno, K., Yokonizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, K., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from Japanese rice  
JOURNAL  
SCIENCE 301 (5631), 376-379 (2003)  
MEDLINE  
22752273  
PUBMED  
12869764  
2 (bases 1 to 1934)  
AUTHORS  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Iehikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Nariyawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ohtsuki, K., Ota, M., Otsuka, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yanagi, M., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.  
Direct Submission  
TITLE  
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@ias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)  
COMMENT  
This clone is one of the 28K full-length cDNA clones from japonica rice.  
URL : <http://cdna01.dna.affrc.go.jp/cdna/>  
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,

Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,  
 Ishikawa, M., Yamada, H., Hotta, I., Kojima, K., Mamiki, T.,  
 Ohneda, E., Yahagi, W., Suzuki, K., Ii, C., Ohtsuki, K., Shishiki, T. and  
 Yamamoto, M.

PAIS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y., Fujiwara, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nariakawa, R., Niihara, J., Oka, M., Ryu, K., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imokani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Masuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Satoh, H., Sakai, K., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES Location/Qualifiers  
xasomtsur, a. and nayashizaki

```
1. .1934
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
clone="J013116M22"
```

## ORIGIN

[illegible]





Db 826 TCGGCGCGGAGCGATGCTCAAGGAGAGGATCAAGTCTGTCACGGTTCGTGTCAACG 885  
QY 797 TCAATCGCGGAGGAGCGGATGCTGAGAGTTCGGGCGAGCGGCAAAACAGGAGAGA 856  
Db 886 TCAATCGGTGACAGGTTCGAGGAGTCTCAACACCAAGGACACACGACATGATTCGAG-- 943  
QY 857 TGAAGCACGACATCTGTCAAGGTTTCATCGAGTTCGGGCGAGCGCGGACGACGCGCGG 916  
Db 944 ----CGAGGATATCTGACAGATTCAT-----CCAGGCAACGACTACGATTCGT 990  
QY 917 GCTTCGGGAGCGATTAAGAGCTTCGGGAGCGTGTGCTCAACTTCGTGATCGCGCGCGGG 976  
Db 991 GGACGTTGATTAAGTACCTGAGAGACATCATATTGAACATTCATAGCCGCAAG 1050  
QY 977 ACACGACGCGAGCGCTGTGCTGTTTCACGACATGCGCATGTCACCGCGGACGTTGG 1036  
Db 1051 ACACGACGCGGCTGTGCTGTTTCCTGTATGATGTGCAACACCGCGGAAGTAC 1110  
QY 1037 CCGAGAGCTGCGCGGAGCTGTGCGGCTTCGAGCGGAGCGCGCGCGGAGGCGG 1096  
Db 1111 AGGAGAGATTCGACCA----- 1129  
QY 1097 TCACGCTGCTGTGCGGCGCGCTGTGACGCGAGCGACAGCGCTTCGCGCGCGCTGG 1156  
Db 1130 -----AGCATGAGGCGCACCAACGCGCGGAGCGCGCTTCATCG 1170  
QY 1157 CGAGTTCGCGGCGCTCTCACTACGACGAGCTTCGGCAAGTGTCTACCTCCACGCT 1216  
Db 1171 AGAGTTCGAGAGCTGACCGAGGAGCTGCAACAGATGACATTCATTCGACGCTG 1230  
QY 1217 GCGTCACGAGAGCTGCGGCTGTACCGCGCGCTCCCTCAGACCCCAAGGGATCTCTG 1276  
Db 1231 CACTGACGAGAGCTGAGGCTATACCTGAGTTCCTGATTAACAGCAGTCTCT 1290  
QY 1277 AGAGAGAGCTGCGGAGCGGAGAGGTSAGGCGCGGCGGAGTGTGAGTACGTCGTC 1336  
Db 1291 CAGACGATGATGTCACCAAGGATTCACGTCAGCAAGGGGAGACATCGTGTTCATCTC 1350  
QY 1337 CCTACTGATGCGGCGGATGAGTACACTGGGCGCGGCGGAGCTTCGCGCGG 1396  
Db 1351 CCTAGCGATGCGGCGGATGAGAGCTGTGGGCGAAGAGCGCTGAATCTTCGCGCGCT 1410  
QY 1397 AGCGGTGATCAAGAGAGATGCGCGTTCGCCCAACGCGTCCGCTTCAAGTTTCACGCGT 1456  
Db 1411 AACCTGCTGATGAGAGCGCTCTTCAGCAGGAGAGCGCTTCAATTTACAGCTT 1470  
QY 1457 TCCAGGCGGCGGAGGATCTGCTGGCAAGGACTCGGCTACTCGAGATGAAGTGG 1516  
Db 1471 TCCAGGCGGCGGAGGATCTGCTCGGAGGAGTTCGCGTACAGGAGATGAAGTCT 1530  
QY 1517 CGCTGGCGCATCTCTTCGCTTCTACAGCTTCGCGCTGCTGGAGGCGACCG---GTGC 1573  
Db 1531 TCGCGGCGGCTGCTCTGCTTCTGCTCAAGCTGCGGAGCGAAGAGATCATCA 1590  
QY 1574 AGTACCGCATGATGACCATCTCTTCCTATGGCGACGCGCTCAAA 1616  
Db 1591 GCTACCGGACCATGATGATGATCTCTCGCTCGATGATGATGATGATGATGATGAT 1633

RESULT 14  
AK061650  
LOCUS  
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:001-036-C03, full insert sequence.  
AK061650  
ACCESSION  
VERSION AK061650.1 GI:32971668  
KEYWORDS FLI cDNA; oligo-capping.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
AUTHORS

1 The Rice Full-Length cDNA Consortium, National Institute of  
Agrobiological Sciences Rice Full-Length cDNA Project Team:  
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,  
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,  
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,  
Ohsuki, K., Shishiki, T., Foundation of Advancement of International  
Science Genome Sequencing & Analysis Group, Ohtomo, Y., Murakami, K.,  
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,  
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,  
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,  
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,  
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,  
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,  
Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,  
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,  
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Yoshino, M. and Hayashizaki, Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from  
japonica rice  
Science 301 (5631), 376-379 (2003)  
2752273

## TITLE

JOURNAL  
MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

2 (bases 1 to 1881)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,  
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,  
Hori, P., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,  
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,  
Kanagawa, S., Kato, H., Kawagashira, N., Kawai, J., Kawamata, M.,  
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,  
Kodama, T., Kojima, K., Kondo, S., Kondo, S., Konno, H., Kouda, M.,  
Koyama, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,  
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,  
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,  
Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,  
Numasaki, R., Ohneda, E., Ohtomo, Y., Ohtsuki, K., Oka, M., Ooka, H.,  
Osato, N., Ota, Y., Ohtsuki, K., Ohtsuki, K., Ohtsuki, K., Sakai, K.,  
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,  
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,  
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,  
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,  
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., and  
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and  
Yoshimura, A.  
Direct Submission

## TITLE

## JOURNAL

## COMMENT

## URL

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of  
Agrobiological Sciences, Department of Molecular Genetics, Head of  
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki  
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,  
Tel:81-29-838-7007, Fax:81-29-838-7007)  
This clone is one of the 28K full-length cDNA clones from japonica  
rice.  
URL: <http://cdna01.dna.affrc.go.jp/cdna/>  
NTAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K.,  
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,  
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and  
Yamamoto, M.  
FAIS Genome Sequencing & Analysis Group: Ohtomo, Y., Iida, Y.,  
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,  
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,  
Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,  
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,  
Yoshimura, A., Matsubara, K. and Murakami, K.  
Genome Exploration Research Group in Riken  
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,  
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
Hiraoka, T., Hori, P., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Kanagawa, S., Kato, H., Kawai, J.,  
Kishikawa-Hirozane, T., Kojima, K., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Oseato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

Location/Qualifiers  
1. 1881  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="001-036-C03"

FEATURES  
source  
1. 1881  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="001-036-C03"

ORIGIN  
Query Match 15.4%; Score 293.8; DB 8; Length 1881;  
Best Local Similarity 53.8; Pred. No. 1.6e-27;  
Matches 813; Conservative 0; Mismatches 597; Indels 101; Gaps 6;  
141 GAAAGGCCCGAGATCATGCGCAGTCATCGCGCAACGGTGGAGCAGCTGAGGAACCTACCA 200  
114 GAAGAAACGAGGATATCGGCCCGTGGCGGCACCATGTTCCACAGCTGCTCACTTCGG 173  
201 CCGGATGCAAGATGCTTGTGGGTACCTGTTCACGGCACAGACAGTGCACGAT 260  
174 CCGGTTGCTGGAGTACACACGAGAGCTCTCCGCAAGTACCGCACCTTCCGCAATGCTCAC 233  
261 GCCGTTTCACTTCTACACCTACCTGCTGACCGCGGTGAATGTGAGGATGCTCTCAAGAC 320  
234 CCGACCTGCAATACGCTTACACCGTGCAGCGCGGCCAACGTCGAGCACATCTCTCAAGAC 293  
321 TAACTTCAACAAATTACCCCAAGGAATCGTGTGATCAGATCCTACATGGAAGTGTCTCTCG 380  
294 CAACTTCGCCAATACGGCAAGGCGCGATGACCCACGAGTGTGGAGAGACCTCTCTCG 353  
381 TGAGGGATTTCAACGCCGACGCGGAGCTGTGGAGGAACGAGGAAGACGCGAGTTT 440  
354 CGACGGGATCTTCAAGTGTGACGCGGCGCATGTGGCGGACGACGCAAGTTCGCCAGCT 413  
441 CGAGTTTGGCTTCAAGAACCTGAGGATTTTCAGGCGCATGTTTTCAGAGAGTACTCCCT 500  
414 CGAGTTTCTCACCCGTTGTGTCGGGACTACACGACGCGCGTGTTCGCGACACCGCCGC 473  
501 GAAGCTGTGGGTATATCTGA-----GCCAGGCATCCAGGACAGGCAAGTGTCTGACAT 554  
474 GGAGCTCGCGGCATCTTGAGCGTGTTCGCGGCGGAGGCGGAGGAGGAGGAGTGGACAT 533  
555 GCAGGAATTTATCATGAGGATGACGCTGAGTCCATCTGCAAGGTTGGGTTGGGTCGA 614  
534 GCAGGATCTGCTGATGGGCGACGCTGGACTCTTTCTTCAGGGTTGGTTTCGGGGTCAA 593  
615 GATCGGCACGCTGTGCCAGATCTCCCGGAGACAGC---TTGCGCGCGGCTTCGATGC 671  
594 CCTTGGCGTCTCTCGGATCCAGCAAGAGGCGCTTGGTGTTCGCGGCGGCTTCGACGA 653  
672 CGCCAAATCATCATCAGCGTGGTTCATCGACCGCGTGTGGCGCATCAAGAGGTTCTT 731  
654 CGCGAGCGGACGAGTGTGTTCCGATTTCTTCGACCTGCTCTGGAAGTTCAGAGGTTCT 713  
732 CCAGTTCGGGTGAGAGCCCTCTAGCGGACGACATCAAGCTGTGGAGAGTTCACCTA 791  
714 CAACATCTCTCGGAGGCAACCATGAGCAGTGCATCCGACCATCAACGACTTCTGTGA 773  
792 CAGCGTGTATCCGCGGAGGAAGCCAGATCTGTCAGGCTCGGCGCGCGGCGGCAACAGA 851  
774 CTCCTATCTCAGAGA-----AGATCGAACAGATGAGCAGAGACCAACGAAAT 824  
852 GAAGATGAAGCAGCATCTGTCTACGTTTCATCGAGCTGGGCGAGGCGGCGGACGAGCG 911  
825 CGCCAAAGAAAGAGACATATGTCGAGGTTTCTGCTCGAGAGGAG-----AAGA 875  
912 CGGCGGCTTCGGGACGATAGAGCCTCGGAGCGTGGTCTCACTTCGTGATCGCGCG 971

876 TCCCGCGCTGCTTCGACAAAGTACATCAGGGAGCATCACTCACTTCGTCATFCGCGG 935  
972 GCGGGACAGGACGCGGACGACGCTGTCTGGTTTCACGCACATGGCCATGTCCACCGCGA 1031  
936 CCGGACAGCAGCGCGGGGACGCTGTCTGGTTTCCTTACGCGCTCTGCAAGAACCGAGCG 995  
1032 CGTGGCCGAGAGCTGCGCGGAGCTGTGGGTTTCGAGGGGAGGCGGCGGCGGAGGA 1091  
996 CGTACAGGACAAAGATCGAGGAGTGGCGGACGCGCACACCGCGGACCGCGACCTCGG 1055  
1092 GGGCGTCAAGCTGCTCTGCGGGCGCTGACGCGGACGACGAGCGTTCGCGCGCGG 1151  
1056 CTTCCAGGATTTCTCTTCTGACAGAGCGCATCAACAGATGCAATGACCTACA 1115  
1152 CFTGGCGAGTTTCGCGGGCTCTCTACCTACGACAGCTTCGCGCAA----- 1196  
1116 CGCAGGTTTGAAGGAGACGCTCGGTTTACCTCGCGTTCCTCTGCTGAGTACTGTTCT 1175  
1197 ----- 1209  
1176 TGGTCTTTTTTTTTTTCGAATAAGAGGAGGAGGAGCTGCTACTGTTGTTGTTGATT 1235  
1210 CAGCGCTGCGTCAACGAGACGCTCGCGCTGTACCCGCGCTCCCT-----CAG 1257  
1236 TTTTCTGAGGTTAGAACTGACGCGCATTTGTTAGTGGCTGTGTTGGCTGTAG 1295  
1258 GACCCCAAGGGATCTCGAGGACGACGCTGTCTCGGACGCGGACGAGTGTAGGCGCGG 1317  
1296 GATGTCAATACTGCTTCTCGGATGGCAGCTTTCGCGGACGCGGACGCGGTGAAGAA 1355  
1318 GGGATGCTCAGCTACGTGCGCTACTCGATGGGCGGAGTGAATCACTGGGCGGCGGAC 1377  
1356 GACATGTTGAATACCAACGCTACCCATGGGCGAGGATGAAGTTCTGTGGGCGGACG 1415  
1378 GCGGCGAGCTTCCGCGCGGAGCGTGGATCAACAGGAGGATGGCGCTTCGCAACGCTCG 1437  
1416 GCGGAGGAGTTCAAGCGCGGAGCGTGGCTCGATGACAGTGGTATGTTCTGCCGAGAGC 1475  
1438 CGTTCAAGTTCAAGCGGTTCCAGCGGCGGCGGAGTCTGCTGGGCAAGACTTCGCGG 1497  
1476 CCGTTCAAGTTCAAGCGGTTCCAGCGGCGGCGGCAAGTCTGCTGGGCAAGAGTTGCG 1535  
1498 TACCTGCGAGTGAAGTGGCGTGGCGGCTTCCTTCGCTTCTAGCTTCGCGGCTGCTG 1557  
1536 TACGCGAGTGAAGTCTTTCGCGCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1595  
1558 GAGGCGGCGACCC 1568  
1596 GATGACGAGCG 1606

## RESULT 15

AK069029

LOCUS

DEFINITION

Oryza sativa (japonica cultivar-group) cDNA clone: J023003E06, full

insert sequence.

AK069029

AK069029.1 GI:32979053

FLI CDNA; CAP trapper.

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzae; Oryza.

REFERENCE

AUTHORS

The Rice Full-Length cDNA Consortium, National Institute of

Agrobiological Sciences Rice Full-Length cDNA Project Team;

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,

Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,

Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,

Ohtsuki, K., Shishiki, T., Foundation of Advancement of International

Science Genome Sequencing &amp; Analysis Group, Otsu, Y., Murakami, K.,

Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:25:24 ; Search time 45 Seconds

(without alignments)

4122.771 Million cell updates/sec

Title: US-10-021-657-2

Perfect score: 3079

Sequence: 1 EFGTREALTPATPSPFPPL.....ICWACKGKRWVSLVWLKP 588

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2816	91.5	543	10 Q94KE5	Q94KE5 zea mays (m
2	2426.5	78.8	544	10 Q8S7V0	Q8S7V0 oryza sativ
3	1894	61.5	524	10 Q9C788	Q9C788 arabidopsis
4	1771	57.5	478	10 Q8GWJ5	Q8GWJ5 arabidopsis
5	1084	35.2	506	10 Q7XTY4	Q7XTY4 oryza sativ
6	1074.5	34.9	515	10 Q8S7S6	Q8S7S6 oryza sativ
7	1040.5	33.8	495	10 Q8S7S4	Q8S7S4 oryza sativ
8	999.5	32.5	560	10 Q8W372	Q8W372 oryza sativ
9	939.5	32.5	560	10 Q7XD91	Q7XD91 oryza sativ
10	936.5	32.4	545	10 Q7XUR3	Q7XUR3 oryza sativ
11	934.5	32.3	553	10 Q8S9L1	Q8S9L1 arabidopsis
12	988	32.1	516	10 Q8S7S7	Q8S7S7 oryza sativ
13	977.5	31.7	511	10 Q64G31	Q64G31 arabidopsis
14	976	31.7	541	10 Q9FVW8	Q9FVW8 triticum ae
15	971.5	31.6	537	10 Q80823	Q80823 arabidopsis
16	970	31.5	554	10 Q9LMM1	Q9LMM1 arabidopsis

17	970	31.5	554	10 Q8GYT6	Q8GYT6 arabidopsis
18	969.5	31.5	523	10 Q9CAD6	Q9CAD6 arabidopsis
19	961.5	31.2	523	10 Q8LAO9	Q8LAO9 arabidopsis
20	913	29.7	559	10 Q9FMV1	Q9FMV1 arabidopsis
21	908.5	29.5	404	10 Q8H091	Q8H091 oryza sativ
22	907.5	29.5	490	10 Q22162	Q22162 arabidopsis
23	904.5	29.4	550	10 Q9LEY2	Q9LEY2 arabidopsis
24	893.5	29.0	488	10 Q8L7B4	Q8L7B4 arabidopsis
25	879.5	28.6	499	10 Q9LXX7	Q9LXX7 arabidopsis
26	877.5	28.5	513	10 Q8XG36	Q8XG36 arabidopsis
27	874.5	28.4	523	10 Q8H810	Q8H810 oryza sativ
28	870.5	28.3	498	10 Q5LNL3	Q5LNL3 arabidopsis
29	867.5	28.2	511	10 Q8LRA0	Q8LRA0 oryza sativ
30	860.5	27.9	512	10 Q8LRA6	Q8LRA6 oryza sativ
31	852.5	27.7	403	10 Q9SNS7	Q9SNS7 arabidopsis
32	847	27.5	508	10 Q8LGH8	Q8LGH8 arabidopsis
33	846	27.5	511	10 Q8LRA4	Q8LRA4 oryza sativ
34	841.5	27.3	508	10 Q8LRA7	Q8LRA7 oryza sativ
35	841	27.3	510	10 Q9FMV7	Q9FMV7 arabidopsis
36	840	27.3	512	10 Q8W2N3	Q8W2N3 vicia sativ
37	839	27.2	510	10 Q8GWY4	Q8GWY4 arabidopsis
38	837	27.2	522	10 Q9FYLO	Q9FYLO arabidopsis
39	832	27.0	511	10 Q8W2N1	Q8W2N1 arabidopsis
40	831.5	27.0	510	10 Q8W2N4	Q8W2N4 nicotiana t
41	823	26.7	506	10 Q9SMP5	Q9SMP5 arabidopsis
42	813.5	26.4	541	10 Q9LTN1	Q9LTN1 arabidopsis
43	802.5	26.1	511	10 Q8W2N2	Q8W2N2 nicotiana t
44	797.5	25.9	496	10 Q9S833	Q9S833 arabidopsis
45	797.5	25.9	524	10 Q94DD8	Q94DD8 oryza sativ

## ALIGNMENTS

### RESULT 1

Q94KE5 PRELIMINARY; PRT; 543 AA.  
ID Q94KE5  
AC Q94KE5  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Cytochrome P450-like protein.  
GN MS-SB200.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fox T.W., Trimnell M.R., Albertsen M.C.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL: AF166297; AAKS2956.1;  
DR GO: GO:0004497; F:monooxygenase activity; IEA.  
DR GO: GO:0006118; P:electron transport; IEA.  
DR InterPro: IPR001128; Cytochrome\_P450.  
DR Pfam: PF00067; P450; 1.  
DR PRINTS: PR00385; P450.  
KW Heme; Monooxygenase; Oxidoreductase.  
SQ SEQUENCE 543 AA; 63308 MW; DB5C3B9D3183CF25 CRC64;

Query Match 91.5%; Score 2816; DB 10; Length 543;  
Best Local Similarity 100.0%; Pred. No. 5.3e-225;  
Matches 541; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 EAHLTATPSPFPPLAGPHKYIALLVLSWILVQWSLRKQKQPRSWPVGATVEQLRN 65  
Db 3 EAHLTATPSPFPPLAGPHKYIALLVLSWILVQWSLRKQKQPRSWPVGATVEQLRN 62  
QY 66 YHRMHWLVGYSRHRVTVDMEFTSYTYIADPVNVEHLKTNFTNYPKGIYRSMDVL 125  
Db 63 YHRMHWLVGYSRHRVTVDMEFTSYTYIADPVNVEHLKTNFTNYPKGIYRSMDVL 122

QY 126 LGDGFNADGELMRKQKRTASFEFASKNLRDFAIIVFREYSLKLSGILSOASKAGKVDM 185  
 Db 123 LGDGFNADGELMRKQKRTASFEFASKNLRDFAIIVFREYSLKLSGILSOASKAGKVDM 182  
 QY 186 QELYMRTLDSICKVGFVEIGTILSPDLSPNSFAQAFDAANIITLRFIDPLWRIKRFPH 245  
 Db 183 QELYMRTLDSICKVGFVEIGTILSPDLSPNSFAQAFDAANIITLRFIDPLWRIKRFPH 242  
 QY 246 VGSALLAQSIKLVDEFTYVIRRRKAEIIVFRASGKQKMKHDIILSRFIELGEAGDGG 305  
 Db 243 VGSALLAQSIKLVDEFTYVIRRRKAEIIVFRASGKQKMKHDIILSRFIELGEAGDGG 302  
 QY 306 GFGDDKSLRDVNLNFVIAGDRTTATLTSWFTWAMSHPDVAEKLRELCAFEAREEG 365  
 Db 303 GFGDDKSLRDVNLNFVIAGDRTTATLTSWFTWAMSHPDVAEKLRELCAFEAREEG 362  
 QY 366 VTLVLCGGADDDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVQDPKGL 425  
 Db 363 VTLVLCGGADDDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVQDPKGL 422  
 QY 426 EDDVLPDGTKVRAGGWTVPYPSMGRMEYNWGDPAASFRPERWINEGDAFRNASPFKFTA 485  
 Db 423 EDDVLPDGTKVRAGGWTVPYPSMGRMEYNWGDPAASFRPERWINEGDAFRNASPFKFTA 482  
 QY 486 FOAGPRICLGKDSAYLQMKMALILFRFYSFRLEGGHPVQYRMWMTILSMHGLKVR 545  
 Db 483 FOAGPRICLGKDSAYLQMKMALILFRFYSFRLEGGHPVQYRMWMTILSMHGLKVR 542  
 QY 546 V 546  
 Db 543 V 543

## RESULT 2

Q8S7V0 ID Q8S7V0 PRELIMINARY; PRT; 544 AA.  
 AC Q8S7V0;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Putative plant cytochrome P-450 protein.  
 OS OSJNB0091P11.12.  
 GN Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,  
 RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,  
 RA Tsitrin T., Riggs F., Hsiao J., Ziemann V., Blunt S., Pai G.,  
 RA VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,  
 RA Salzberg S.L., White O., Fraser C.M.;  
 RT "Oryza sativa chromosome 3 BAC OSJNB0091P11 genomic sequence.";  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 DR EMBL; AC073556; AAL84318.1;  
 DR Gramene; Q8S7V0;  
 DR GO; GO:0004497; F:monooxygenase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 KW Heme; Monooxygenase; Oxidoreductase.  
 SQ SEQUENCE 544 AA; 61182 MW; 1A52011B956F92FD CRC64;

## Query Match

Best Local Similarity 78.8%; Score 2426.5; DB 10; Length 544;  
 Matches 471; Conservative 32; Mismatches 31; Indels 11; Gaps 5;

QY 6 EAHLTATPSPFPPLAGPHKYIALLLVLSWLVORWSLRKQKPPRSWPVIGATVEQLRN 65  
 Db 7 EAHLPVMT--SFPFVAGIHKLIATFLVLSWLVORWSLRKQKPPRSWPVIGATVEQLKN 64  
 QY 66 YRKHMDLVGLYLSRHRTVTVDMPFTSYTYIADPVNVVHVKTNFTNPKGIVVRSYMDVL 125  
 Db 65 YRKHMDLVGLYLSRHRTVTVDMPFTSYTYIADPVNVVHVKTNFTNPKGIVVRSYMDVL 124  
 QY 126 LGDGFNADGELMRKQKRTASFEFASKNLRDFAIIVFREYSLKLSGILSOASKAGKVDM 185  
 Db 125 LGDGFNADGELMRKQKRTASFEFASKNLRDFAIIVFREYSLKLSGILSOASKAGKVDM 184  
 QY 186 QELYMRTLDSICKVGFVEIGTILSPDLSPNSFAQAFDAANIITLRFIDPLWRIKRFPH 245  
 Db 185 QELYMRTLDSICKVGFVEIGTILSPDLSPNSFAQAFDAANIITLRFIDPLWRIKRFPH 244  
 QY 246 VGSALLAQSIKLVDEFTYVIRRRKAEIIVFRASGKQKMKHDIILSRFIELGEAGD-- 303  
 Db 245 VGSALLAQSIKLVDEFTYVIRRRKAEIIVFRASGKQKMKHDIILSRFIELGEAGD-- 304  
 QY 304 GFGDDKSLRDVNLNFVIAGDRTTATLTSWFTWAMSHPDVAEKLRELCAFEAREARE 363  
 Db 305 GFGDDKSLRDVNLNFVIAGDRTTATLTSWFTWAMSHPDVAEKLRELCAFEAREARE 364  
 QY 364 EGVTLVLCGGADDDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVQDPK 422  
 Db 365 EGVTLVLCGGADDDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVQDPK 419  
 QY 423 GLEDDVLPDGTKVRAGGWTVPYPSMGRMEYNWGDPAASFRPERWINEGDAFRNASPF 481  
 Db 420 GLEDDVLPDGTKVRAGGWTVPYPSMGRMEYNWGDPAASFRPERWINEGDAFRNASPF 479  
 QY 482 KFTAFQAGPRICLGKDSAYLQMKMALILFRFYSFRLEGGHPVQYRMWMTILSMHGLKVR 541  
 Db 480 KFTAFQAGPRICLGKDSAYLQMKMALILFRFYSFRLEGGHPVQYRMWMTILSMHGLKVR 539  
 QY 542 VSRV 546  
 Db 540 VSTSV 544

RESULT 3  
 Q9C788 ID Q9C788 PRELIMINARY; PRT; 524 AA.  
 AC Q9C788;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DE Cytochrome P450, putative.  
 DE Cytochrome P450, putative.  
 GN F1013\_15.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., S.Y.,  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
 RA Militschker J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,









```

QY 375 DADDKFAARVAQFAGLLTYDSLGLKYLHACVTTETLRILYPAVDPQPKGILEDDVLPDGT 434
Db 397 DEDDLVPH-----EEVKMEYLHAALSEALRSLYPSVVDHKEVVEDEVFPDGT 445
QY 435 KVRAGGVTYVYPSMGMEYNWGPDAASFRPERWINEGDAFRNASPKFTAFQAGPRICL 494
Db 446 VLKKGTKVIYAMYTMRMESINGEDCREYKPERWL-RDGRFMGESAYKFTAFNGGPRICL 504
QY 495 GKDSAYLQMKALAILFRFYSLLEGGHPVQYRMMTILSMAHGLKVRVSR 544
Db 505 GKDFAYYQMKFAAASILRRYHVVDGHPVAPKQVALTMYMKHGLKVKLT 554

RESULT 9
Q7XD91
ID Q7XD91 PRELIMINARY; PRT; 560 AA.
AC Q7XD91
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Putative cytochrome P450 protein.
GN OSNBA0029C15.3.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RL chromosome 10.";
RN [2]
Science 300:1566-1569(2003).
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE017107; AAP54351.1;
SQ SEQUENCE 560 AA; 61698 MW; 0291968D176CC384 CRC64;

```

```

Query Match 32.5%; Score 999.5; DB 10; Length 560;
Best Local Similarity 40.6%; Pred. No. 4.3e-74;
Matches 215; Conservative 91; Mismatches 193; Indels 31; Gaps 10;

QY 27 IALLVVLVILVQWLSLKKQKPRSWPVIGATVEQLRNH-RMHDWLGYL-SRRHVT 84
Db 44 VELLVAVSIFVAHSQRSSQGLPSWPLVGLPSLLGLRGDMYEWLTVLASRGGTFT 103
QY 85 VDMPTSTYTY---IADPVNVEHLKTNFTNPKGIYVRSYMDVLLGDGIFNADGELWRQ 141
Db 104 FHGPWLTNLHCVVTSDPRLNHLTKTQSGFKGPFYFRTVRLDGLDGFADDEVWRQ 163
QY 142 RTAFGEFASKNLRFSA-----IVFREYSLKSLGISQASKAGKVDVDMQELMYMTLDS 196
Db 164 RKAASLEFSAFRALTASSLVELVH-----RLRLVGLDAEAGDAVDLQDVLRLTFDN 219
QY 197 ICKVGFGEIGTSLDLPENSFAQAFDAANIITLRFIDP--LWRIKRFHVGSEALLAQ 254
Db 220 VCMIAFGVDPGCLRPLPIPAKAFEDATEATIVRFVPTAVWRAMRALGVGHERVLQR 279
QY 255 SIKLVDEFTYSVIRRKAEIIVEVRASGQKQMKHDLISFELGEAGDGGGFGDDKSLR 314
Db 280 SLAGVDREAYDVTRQKEV--AAGGGGGGGRSLLTFTKMRDA-DTGAAYASDKFLR 336
QY 315 DVILNFTVAGRTDTATTLTSWFTMHMSHEDVAEKLRELCAFEAREAREEGVTLVLCGA 374
Db 337 DICVNFILAGRTSSVALAWEFWLLNKNPVAEAKILEEDDIIAARRSPAPAVANGA 396
QY 375 DADDKFAARVAQFAGLLTYDSLGLKYLHACVTTETLRILYPAVDPQPKGILEDDVLPDGT 434

```

```

Db 397 DEDDLVPH-----EEVKMEYLHAALSEALRSLYPSVVDHKEVVEDEVFPDGT 445
QY 435 KVRAGGVTYVYPSMGMEYNWGPDAASFRPERWINEGDAFRNASPKFTAFQAGPRICL 494
Db 446 VLKKGTKVIYAMYTMRMESINGEDCREYKPERWL-RDGRFMGESAYKFTAFNGGPRICL 504
QY 495 GKDSAYLQMKALAILFRFYSLLEGGHPVQYRMMTILSMAHGLKVRVSR 544
Db 505 GKDFAYYQMKFAAASILRRYHVVDGHPVAPKQVALTMYMKHGLKVKLT 554

RESULT 10
Q7XUR3
ID Q7XUR3 PRELIMINARY; PRT; 545 AA.
AC Q7XUR3
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE OSJNB0084K11.7 protein.
GN OSJNB0084K11.7.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Lu Y.Q., Xu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying X., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL606887; CAD41073.1;
SQ SEQUENCE 545 AA; 60100 MW; 8BE20C4D4AB66A79 CRC64;

```

```

Query Match 32.4%; Score 996.5; DB 10; Length 545;
Best Local Similarity 42.0%; Pred. No. 7.2e-74;
Matches 226; Conservative 77; Mismatches 182; Indels 53; Gaps 14;

QY 27 IALLVVLVILVQWLSLKKQKPRSWPVIGATVEQLRNH-RMHDWLGYL-SRRHVT 81
Db 10 VAAVAAYMAWF-----WRMSRGLSGRVPVVGSLPGLVHAEDMEHTAANLRTRGTQ 65
QY 82 TVTVDMPTTS-----YTIADPVNVEHLKTNFTNPKGIYVRSYMDVLLGDGIFNADGE 136
Db 66 TCFVAVGLARGGILVTCTDPRLNHLKSRFDNPKGPFVGHGVGDLGDFGPNSDGE 125
QY 137 LWRKORTASPEFASKNLRD-FSAIVFREYSLKSLGISQASKA-----GKVDVDMQELYM 190
Db 126 TWVAQRKTALEFTTTLRTAMSRKVSISIRLLPILSDAAAGGGGGATVDLQDULL 185
QY 191 RMTLDSICKVGFGEIGTSLDLPENSFAQAFDAANIITLRFIDP--LWRIKRFHVGVS 248
Db 186 RLTFDNLGLAFGKDPETLARGLPENDFASAFDRATEATLNRFPFECVWFKKWMLGM 245
QY 249 EALLAQSTIKLVDEFTYSVIRRKAEIIVEVRASGQKQEM--KHDILSRFIEIGEADGGG 306
Db 246 ETTLARSVQHVDRYLSAVIKARKLELAGNGKGDASSATPHDDLLSRFMKGTYSDE--- 302
QY 307 FGDDKSLRDVVLNFTVAGRTDTATTLTSWFTMHMSHEDVAEKLRELCAFEAREAREEGV 366
Db 303 -----SLCHVALNFILAGRTSSVALSWFFWVLTSTHFAVERKIVRELCT----- 346
QY 367 TLVLGGGADDDKFAARVAQFAGLLTYDSLGLKYLHACVTTETLRILYPAVDPQPKGILE 426
Db 347 --VLAASRGADDPAL-----WLAAPLNFEELDQLVYLKALSETLRLYPSVPEDSKHVA 399
QY 427 DDVLDPDGTKVRAGGVTVYVPSMGMEYNWGPDAASFRPERWINEGDA-FRNASPFKFTA 485

```





us-10-021-657-2.rspt

Mon Mar 8 10:13:28 2004

DT	01-MAR-2001 (TREMblrel. 16, Created)	RESULT 15
DT	01-MAR-2001 (TREMblrel. 16, Last sequence update)	OB0823
DT	01-JUN-2003 (TREMblrel. 24, Last annotation update)	ID
DT	Cytochrome P450.	PRELIMINARY;
OS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	PRT; 537 AA.
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;	
OC	Triticeae; Triticum.	
OC	NCBI_TaxID=4565;	
RN	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	
RC	MEDLINE=20322901; PubMed=10864454; Schalk M., Pallett K., Schaller H.,	
RX	Batard Y., Henn A., Nedelkina S., Schalk M., Pallett K., Schaller H.,	
RA	Werk-Reichardt D.;	
RA	"Increasing expression of P450 and P450-reductase proteins from	
RT	monocots in heterologous systems."	
RT	Arch. Biochem. Biophys. 379:161-169(2000).	
RL	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.	
CC	EMBL; AF123609; AAG17470.1; -	
DR	GO; GO:0004497; F:monooxygenase activity; IEA.	
DR	GO; GO:0006118; P:electron transport; IEA.	
DR	InterPro; IPR001128; Cytochrome_P450.	
DR	Pfam; PF00067; P450; 1.	
DR	PRINTS; PR00385; P450.	
DR	PROSITE; PS00086; Cytochrome_P450; 1.	
KW	Heme; Monooxygenase; Oxidoreductase.	
QW	SEQUENCE 541 AA; 59871 MW; DDAC8B8C90861C00 CRC64;	
SQ	SEQUENCE 541 AA; 59871 MW; DDAC8B8C90861C00 CRC64;	
Query Match	31.7%; Score 976; DB 10; Length 541;	
Best Local Similarity	39.9%; Pred. No. 3.6e-72;	
Matches	222; Conservative	
QY	27 IALLVLSNLVQWNSL-RKQGPSPWVTAIVQELNVRHMDMLVGLYSR-----HR 81	
Db	10 VSAVAAMWFF-----WRSRGLGRFVWVFLGSLGVLQVQHAEDHWEIAGNLRAGGTGY 65	
QY	82 TVTVMDPFTS-----YTVIADPNVNHVLTNTNPKGIVVRSYMDVLLGGIENADGE 136	
Db	66 TCIFAVEGVARGGLVTTCDPNLEHLKAREFDNPKGPFVGHVPRDLGGIFNSDGD 125	
QY	137 LWRKORTASFEFASKNLRD-PSAIVFREYSLKIGLSOASVAGKVVDMDQELYMRTLD 195	
Db	126 TWLAQRKTALEFTRTLRTAMSRWSESITHGRLPLPILADAAGKAQVLDLRLITFD 185	
QY	196 SICKVGFEVGTSPDLSPNSFAQAFDAANIITLREIDP--LWRIKRFHVSEALLA 253	
Db	186 NICGLAFGKDPETLQAGLPENEFASAFDRATEATLNRIFFELWCKKGLGMMETTLT 245	
QY	254 QIKLVDEFTYVIRRRKAEIVFVRSAGKQKMKHDILSRFELGEAGDGGGFGDKSL 313	
Db	246 SSNAHVQYLAIVIKRKLKLELAGNGKCDTAAHDDLLSRFRKGSYSDE-----SL 297	
QY	314 RDVVLNFVAGRTTATTLSWETHMAMSHPDVAEKRLRELCAFEAREEAGVTVLVLGG 373	
Db	298 QHVALNFILAGRTSSVALSWFFWVSTHFAVERKIVRELCS-----VLAAS 344	
QY	374 ADADKFAFAVAQFAGLLTYDSLGKLVYLHACVETTLRLYPVAPQDPKLEDDVLPDG 433	
Db	345 RGAHDPAL-----WLAFFTEELDRVLVLAASETLRLYSPVDSKHYVADYLPDG 399	
QY	434 TKVAGGMVTVYVYSGMGRMYNWDPAASERPERWINEGDA-FRNASFKTTAFQAGPRI 492	
Db	400 TTVPAGSVTVYSISAGRMKGVNGDCLEFRERWLSADGTKEQHDYSYKFAVAGPRV 459	
QY	493 CLGKDSAYLQMK-MALAILFRYSFRLLEGHPCVQYRMNTILSMAHLKVRVSRVACHGL 551	
Db	460 CLGKDLAYLQMKNIAGSVLLR-HRLITVAPGRVQEKMSLTLFWKGLRMEV----- 509	
QY	552 DMDIVLPNPRQITVL 567	
Db	510 -----RPRDLAPVL 518	

Search completed: March 2, 2004, 15:29:11  
Job time : 47 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:22:09 ; Search time 18 seconds  
(without alignments)  
1700.958 Million cell updates/sec

Title: US-10-021-657-2

Perfect score: 3079  
Sequence: 1 EFGTEAHLTPATPSFFPL.....ICMHACKGRWVSLVWLKP 588

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	994.5	32.3	553	1 C862 ARATH	O23066 arabidopsis
2	880.5	28.6	513	1 C861 ARATH	P48422 arabidopsis
3	818.5	26.6	513	1 C942 VICSA	P98188 vicia sativ
4	785.5	25.5	514	1 C941 VICSA	O81117 vicia sativ
5	427	13.9	519	1 CP5V CANAP	P43083 candida api
6	423	13.7	519	1 CP5L DEBHA	O9V757 debaryomyce
7	411	13.3	519	1 CP5W CANAP	Q12573 candida api
8	406.5	13.2	523	1 CP5C CANNA	P16496 candida mal
9	402	13.1	524	1 CP5F CANTR	P30608 candida tro
10	400.5	13.0	523	1 CP5E CANNA	P24458 candida mal
11	398.5	12.9	531	1 C4E1 DROME	O9V4t5 drosophila
12	395	12.8	519	1 CP5M CANNA	O9V758 debaryomyce
13	387	12.6	519	1 CPFC HUMAN	Q12589 candida mal
14	386.5	12.6	524	1 CP5D CANNA	Q12582 homo sapien
15	395	12.5	538	1 CP5J CANNA	Q12588 candida mal
16	384.5	12.5	519	1 CP5J CANNA	P16141 candida tro
17	383.5	12.5	520	1 YRV5 CAEEL	P30607 candida tro
18	381	12.4	522	1 CP5B CANTR	Q27519 caenorhabdi
19	380.5	12.4	518	1 YRVA CAEEL	Q46051 drosophila
20	380	12.3	507	1 C4DE DROME	P98187 homo sapien
21	377.5	12.3	526	1 CP5H HUMAN	Q27589 drosophila
22	377	12.2	520	1 CP5H HUMAN	O98477 homo sapien
23	375	12.2	501	1 YS24 CAEEL	Q09653 caenorhabdi
24	372.5	12.1	534	1 YRV8 CAEEL	Q08477 homo sapien
25	371	12.0	520	1 YRV8 CAEEL	P10615 candida tro
26	370.5	12.0	543	1 CP5A CANTR	Q12585 candida mal
27	370	12.0	543	1 CP5T CANNA	P30610 candida tro
28	369.5	12.0	505	1 CP5T CANNA	O12586 candida mal
29	368	12.0	517	1 CP5H CANTR	O48921 glycine max
30	367	11.9	521	1 CP5T CANNA	P48323 homo sapien
31	367	11.9	576	1 C972 SOYBN	P11707 oryctolagus
32	366	11.9	520	1 CP5T CANNA	
33	365	11.9	501	1 CP36 RABIT	

34	362	11.8	503	1 CP39 RAT	P51538 rattus norv
35	360	11.7	526	1 CP55 RAT	P51870 rattus norv
36	359.5	11.7	483	1 C311 DROME	Q9VYq7 drosophila
37	357	11.6	552	1 C971 PEA	Q43078 pisum sativ
38	356.5	11.6	524	1 CP5L RAT	P33274 rattus norv
39	356	11.6	507	1 CP5G CANTR	P30609 candida tro
40	355.5	11.5	518	1 YRV3 CAEEL	Q27515 caenorhabdi
41	355	11.5	520	1 YRV3 CAEEL	Q27514 caenorhabdi
42	354.5	11.5	510	1 C4DK DROME	Q9W011 drosophila
43	353.5	11.5	510	1 C312 DROME	Q9V0N6 drosophila
44	353	11.5	524	1 CP5B HUMAN	Q9bbi6 homo sapien
45	351.5	11.4	504	1 CP3B MOUSE	Q64459 mus musculu

ALIGNMENTS

RESULT 1

C862 ARATH STANDARD; PRT; 553 AA.

AC O23066; (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cytochrome P450 86A2 (EC 1.14.-.-).

GN CYP8A2 OR AT4G0360 OR A.IG005110.21 OR F5110.21.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC euroids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

[1]

RN SEQUENCE FROM N.A.

RP STRAIN=cv. Columbia;

RC Andrews S.;

RA Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

[2]

RN SEQUENCE FROM N.A.

RP STRAIN=cv. Columbia;

RC MEDLINE=20083488; PubMed=10617198;

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,

RA Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,

RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M., Mueller M.,

RA Weichselgartner M., de Simone V., Obermaier P., Watson M., Schmidheini T.,

RA Kreis M., Delseny M., Puigdomenech P., Boutry M., Bancroft I.,

RA Reichert B., Portetelle D., Perez-Alonso M., Wedler H., Ridley P.,

RA Vos P., Hoheisel J., Zimmermann W., Lamm H., Robben J., Vandenbussche F.,

RA Langham S.-A., McCullagh B., Bilham L., Chuang Y.-J., Vandenbussche F.,

RA Van der Schueren J., Grymonprez B., Bastiaens I., Aert R., Defoor E.,

RA Braeken M., Weltjens I., Voet M., Bastiaens I., Hilbert H., Braun M.,

RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Dirkse W.,

RA Holzer E., Brandt A., Peters S., van Staveren M., Koetter P.,

RA Moolenaar P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,

RA Bernier S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,

RA De Keyser A., Buyschaert C., Gielens J., Villarroel R., De Clercq R.,

RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,

RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McClay K., Mayes R.,

RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,

RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,

RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Partmann B., Granderath K., Dauner D., Herzl A.,

RA Neumann S., Argitoni A., Vitale D., Ligouri R., Piravandi E.,

RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,

RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,

RA Chedofor T., Cooke R., Berger C., Vonfort A., Casacuberta E.,

RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,

RA Perez-Perez A., Purnelle B., Bent E., Johnson S., facon D., Jesse T.,

RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,

RA Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,

RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,

RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,

RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,

RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,



Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,  
 Nelson J., Spieth J., Ryan E., Andrews K., Geisel C., Layman D.,  
 Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 Antoniou B., Zidan M., Strong C., Sun H., Lamar B., Jordan C.,  
 Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 Granat S., Shondy N., Rasegawa A., Hameed A., Lodhi M., Johnson A.,  
 Chen E., Marra M.A., Martienssen R., McCombie W.R.;  
 "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 thaliana";  
 Nature 402:769-777(1999).  
 -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DB EMBL; AF013293; AAB62843.1; -;  
 DR EMBL; AF195115; AAF02801.1; -;  
 DR EMBL; AL161471; CAB80794.1; -;  
 DR F01; T01535; T01535;  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 DR OXIDOREDUCTASE; Monooxygenase; Heme; Transmembrane; Multigene family.  
 TRANSMEM 20 POTENTIAL.  
 FT METAL 459 459 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT SEQUENCE 553 AA; 62530 MW; 11802EDC325DBCB6 CRC64;  
 SQ  
 Query Match 32.3%; Score 994.5; DB 1; Length 553;  
 Best Local Similarity 41.0%; Pred. No. 1.2e-56;  
 Matches 226; Conservative 92; Mismatches 182; Indels 51; Gaps 15;  
 29 LLLVLS--WLVQSLRQKQPRSPVITGATVEQLNVRHMDMLVGYL----SRHRT 82  
 8 LLLVVAAYWLFQRIIS--RWLKGVPWPLVLSGLIEQDRDHRDMLTENLRACGGTYQT 66  
 83 VTVDMEPTS-----YTVIADPVNVEHVLKTNFNYPKGIIVRSYMDVLLGDGIFNADGEL 137  
 67 CIGAVPLAKQGLVTVTCDPKNEHMLKTRFDNYPKGPQWQVFDHFLGQGFNSDGD 126  
 138 WRKQRTASPEFASKNLRD--FSATVFREYSIKLSGILSQASKAGKVVDQMELYNRYTLDS 196  
 127 WLFQRTAALEFTRTLRQAMGRVWNRGILKRFCEPILFETAGNNYEPVDQLRLRLTFDN 186  
 197 ICKVGFGEIGTSPDLSPNSFAQAFDAANIIITRLRIDP--LWRIKRFHFVSGEALLAQ 254  
 187 ICGAFAGKDTKTCAPGUPENGAFADPATEASLQRFILPEFLWLLKWLGLGLEVLSLR 246  
 255 SIKLVDEFTYSVIRRRKAEIIVEVRASGKQEKMKH--DILSRFIEGADGGGFGDDKSL 313  
 247 SLGIDGVLDAVINTRKQELLSQRESGVQ-----RHDDLRSRMKKKQDS-----YSETFL 297  
 314 RDVNLNFIAGRTDTATTLSTHFTMAMSHPDVAEKLRELCAFEARESGVTLVLCGG 373  
 298 RHVALNFILAGRTSSVALSFFWLLITHTPTVEDKIVREICS-----VLITRFG 346  
 374 ADADKFAARVAQAFAGLLTVDSLGKLVYLHACVETLRLYPVAVPQDPKGLIEDDVLDPG 433  
 347 TDVS-----SWTAEPLEFDEVDRLVYLKAALETURLIPSVEDSKVYVNDLIDFG 398  
 434 TKVRAGGMVTVYPYSGRMENYNGWPDAAFRPERWIN--EDGAFRNASPFKFTAFQAGPRI 492  
 399 TVFVAGSSVTYSIYAGRMKSTWGEDCLEFKFERWISPDGKFNVDQVRFVAFNAGPRI 458  
 493 CLGKDSAYLQMK-MALAILFRYSPFLLEGHPVQVYRMVTLISMAHGLKQVRSRAVCHGDL 551

Db 459 CLGKDSAYLQMKTIARAALLR-HRLTVAPGHKVEQKMSLTFLMKNGLLVNVHVR-----DL 513  
 QY 552 DMDIVFLNPRQ 562  
 Db 514 EVMKSLVPKE 524  
 RESULT 2  
 C861 ARATH STANDARD; PRT; 513 AA.  
 ID C861 ARATH  
 AC P48422; O9FIM3;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 86A1 (EC 1.14.--) (CYPLXXXVI) (P450-dependent fatty  
 acid omega-hydroxylase).  
 GN CYP86A1 OR CYP86 OR AT5G58860 OR K19W22.6.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Benveniste I., Durst F.;  
 RT "Cloning, sequencing and expression of CYP86, a new cytochrome P450  
 from Arabidopsis thaliana";  
 RL (In) Plant Gene Register PGR95-074.  
 RN [2]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=98162595; PubMed=9500987;  
 RA Benveniste I., Tjiet N., Adas F., Phillips G., Salaun J.P., Durst F.;  
 RT "CYP86A1 from Arabidopsis thaliana encodes a cytochrome P450-dependent  
 fatty acid omega-hydroxylase";  
 RL Biochem. Biophys. Res. Commun. 243:688-693(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=99156233; PubMed=10048488;  
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.  
 Sequence features of the regions of 1,081,958 bp covered by seventeen  
 physically assigned B1 and TAC clones";  
 RL DNA Res. 5:379-391(1998).  
 CC -1- FUNCTION: Catalyzes the omega-hydroxylation of various fatty acids  
 (FA). Acts on saturated and unsaturated fatty acids with chain  
 lengths from C12 to C18 but not on hexadecane.  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DB EMBL; X90458; CAA62082.1; -;  
 DR EMBL; AB016885; BAB09631.1; -;  
 DR PIR; JC5965; JC5965.  
 DR HSP; P14779; LJPZ.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 DR OXIDOREDUCTASE; Monooxygenase; Heme; Transmembrane; Multigene family.  
 TRANSMEM 27 POTENTIAL.  
 FT METAL 456 456 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT CONFLICT 387 388 DD -> EH (IN REF. 1 AND 2).  
 FT SEQUENCE 513 AA; 58553 MW; 441A0CBCEB2DB5 CRC64;  
 SQ

Query Match	28.6%; Score 880.5; DB 1; Length 513;
Best Local Similarity	39.6%; Pred. No. 2.3e-49;
Matches 210; Conservative	84; Mismatches 181; Indels 55; Gaps 13;
QY	27 IALLIVLSIIVQWSLKKQKQPSWPGVIGATVQLRNYHRMHWLVGYL-----SRHRT 82
DB	13 VAALSVALWFY---FLSRLTGPKVLPFVGLSPFLYLIANRSHIHDIADNLRAATGGTVQT 69
QY	83 VTVDMPEFTS-----YTYIADPNVNHVLTNTNPKGIIVYRSYMDVLGDFGNADGBL 137
DB	70 CTWNPFPVAKAGFYVTCHPKANVEHLKTRDNPVKGPWRAAHPHLLGQGFNSDGT 129
QY	138 WRKQRTASFEFASKNLDFSA-IVFREYSLKLSGILSOAKGKVVDMQOELVMRWTLDS 196
DB	130 WLMQRTAALEFTRTRQAMARVWNGTIKNRLWLILDRVQNNKPDVLQDLFLRLTFDN 189
QY	197 ICKVGFVIGLISPLDPENSPAQAADAAN--IIITLRFIDPLWRIKRFHHVGSALLAQ 254
DB	190 ICGLTFGKDPETLSLDPNPFVSVAFDTATEATLKSLTYTGFLWRIQKAWGIGSEDKLKK 249
QY	255 SIKLVDEFTYSVIRRRKABIVEVRASQKQKMDHILSRFIELGEAGDDGGFGDDKSLR 314
DB	250 SLEVVETYMNDALDARK-----NSPDDLLSRFL---KKRDVNGVNLPTDVLQ 294
QY	315 DVVLNFVIAGRDATTATLSWFTHMAMSHDPVAKLRRLRELCFAFAERAREEGVTVLVCGGA 374
DB	295 RIALNFVLAGRDTSSVALSFFWFLVMNRREVETKIVNEL-----SMVLKETR 341
QY	375 DADDKFAARVAQFAGLLTYDSLGKLVYHACVTEITRLYPAVPODPKGTLEDVLPDGT 434
DB	342 GNDQEKWTEPLEF-----DEADRLVYLKAALAEITLRLYPSVPODPKYVDDVLPDGT 395
QY	435 KVRAGGMVTVPYSGRMEYNWGPDAASPRPERWINEGAFRNASP---PKFTAFQAGER 491
DB	396 FVPRGSTVTYSISIGMKTIWGEDCLEFRPERWLTADGE-RFETPKDGYKFAFNAGPR 454
QY	492 ICLGKDSAYLQMK-MALAILFRFYSFKLLEGHFPVQYRMMTILSMAHGLKV 540
DB	455 TCLGKDLAYNCMKSVASAVLLRYRVFP-VPGHRVEQKMSLTLFMKNGLRV 503
RESULT 3	
C942	VICSA
ID	C942 VICSA STANDARD; PRT; 513 AA.
AC	P98188;
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DE	Cytochrome P450 94A2 (EC 1.14.-.-) (P450-dependent fatty acid
DE	omega-hydroxylase).
GN	CYP94A2 OR VAGH811.
OS	Vicia sativa (Spring vetch) (Tare).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC	eucosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.
OC	NCBI_TaxID=3908;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Seedling;
RX	MEDLINE=9935375; PubMed=10405339;
RA	Le Bouquin R., Pinot F., Benveniste I., Salauen J.-P., Durst F.;
RT	"Cloning and functional characterization of CYP94A2, a medium chain
RT	fatty acid hydroxylase from Vicia sativa."
RL	Biochem. Biophys. Res. Commun. 261:156-162(1999).
CC	-!- FUNCTION: Catalyzes the omega-hydroxylation of various fatty acids
CC	(FA). The substrate specificity is higher for myristate > laurate
CC	= palmitate (C14>C16=C12).
CC	-!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC	similarity).
CC	-!- TISSUE SPECIFICITY: Weakly expressed in seedlings.
CC	-!- INDUCTION: By clofibrate; weakly.
CC	-!- SIMILARITY: Belongs to the cytochrome P450 family.
This SWISS-PROT entry is copyright. It is produced through a collaboration	
between the Swiss Institute of Bioinformatics and the EMBL outstation	
at the European Bioinformatics Institute. There are no restrictions on its	
use by non-profit institutions as long as its content is in no way	
modified and this statement is not removed. Usage by and for commercial	
entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
EMBL; AF092917; AAC33645.1; --	
InterPro; IPR001128; Cytochrome_P450.	
Pfam; PF00067; P450; 1.	
PRINTS; PR00385; P450.	
PROSITE; PS00086; CYTOCHROME_P450; FALSE NEG.	
Oxidoreductase; Monooxygenase; Heme; Transmembrane;	
Endoplasmic reticulum; Multigene family.	
TRANSMEM 7 24	
POTENTIAL.	
FT METAL 455 455	
IRON (HEME AXIAL LIGAND) (BY SIMILARITY).	
SEQUENCE 513 AA; 58419 MW; BEA213B45837B1A7 CRC64;	
Query Match	
Best Local Similarity	
Matches 188; Conservative	
QY	30 LLVLSWILVQR---WSL---RKQKG-----PRSWPVIGATVBOELRYHRM 69
DB	3 LETLSWLLPFSLSLFWFLATKTKSKPKTPSTTNTPIKSYPIFGSAFSLANFHR 62
QY	70 HDLVGYLSR--HRTVTVDMPP--TSYTIADPNVNHVLTNTNPKGIIVYRSYMDVL 126
DB	63 IQWTSILOITIPSSITFVLHRPFGARQVTAQPAVQVHILRTNFTCYGKGTTFYQSINDFL 122
QY	127 GGGIFNADGELMRKQKRTASFEFASKNLDF--SAIVFREYSLKLSGILSOAKGKVVDM 185
DB	123 GGGIFNADGESWFKPQKQISHEFNTRSLKRFVTVVDVLSDLVLPVLSQASNSQTTLD 182
QY	186 QELYMNTLDSICKVGFVIGLISPLDPENSPAQAADAANIITLR---FIDPLWRIKR 242
DB	183 QDILQRLTFDNCIAFGYDPEYLLSLPEIPFAKAFDESSQLSIBRLNALIPLLMKVKR 242
QY	243 FFHVGSALLAQSIKLVDEFTYSVIRRRKABIVEVRASQKQKMDHILSRFIELGEAGD 302
DB	243 FLNIGVERQLKEAVERGLATKIVKNKKELKALQSESESV--DLLSRFL----- 293
QY	303 DGGFGDDKSLRDVNLNFVLAGDRTATTATLSWFTHMAMSHDPVAKLRRLRELCFAFAERAR 362
DB	294 -SSGHSDESEFTDMVISIILLAGDRTTSAALTFFWLLSKSHSVENEILKEI-----TGK 346
QY	363 EEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYHACVTEITRLYPAVPODPK 422
DB	347 SETV-----GYDEVKDMVYTHAALCESMRLYPLPLPYDTK 380
QY	423 GILEDDVLPDGTKVRAGMVTYTPYSGRMEYNWGPDAASFRPERWINEO-----GAFRNA 478
DB	381 VAVHDDVLPDGTLVKKGWRVYTHIYAMGRSEKINGPDWAFRPERWLSRDEVGKWSFVGI 440
QY	479 SPFXTAFAQAGPRICLKDSAYLQMKWALA--ILFRFYSF--RLLEGHFPVQYRMMTILSMAH 536
DB	441 DYISYFVFOAGPRVICIGREMAFLQMKRVACIMGRFRVVPANVVGIEPEYTAHTFTSVMKG 500
QY	537 GLKVRVSR 544
DB	501 GFPVKIEK 508
RESULT 4	
C941	VICSA
ID	C941 VICSA STANDARD; PRT; 514 AA.
AC	O81117;
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Cytochrome P450 94A1 (EC 1.14.-.-) (P450-dependent fatty acid



QY	231	LRFI-----DPLWRIKFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIWEVRAS	280
Db	238	VRVIVGAARFLIWTPEKRSCKVCH-----NFIDYFYKAL-----AT	275
QY	281	GKQKXKHDLISRFTELGEAGDGGFGDKSLRDVLNLFVAGRDTTATTLISWETHMAM	340
Db	276	FMKQEQEAD---RYFIRELTKE---TSDPRVIRQALNILLAGRDTTAGLLSFITYILG	329
QY	341	SHPDVAELRLRELCAFEARAREEGVTILVLCGGADADKAFARVAQFAGLLTYDSLGLK	400
Db	330	AYPEVAELRE-----AVLSERGSTDVETP-----TFEQLKQC	362
QY	401	VYLHACVTTLRLYPVPODPKGILEDDVLDPG-----TKVRAGGMVTYVPYSGMYX	452
Db	363	KVLQNVIREVLRLHFNVPINFRQAIVDTKLPTGGGNGDQPVFPVPGQNVFYSTYSNQRR	422
QY	453	EYNWGDPAASFERPERWINEGDAFRNASPFKFTAFQAGPRICLKDSAYLQMKMALILFR	512
Db	423	TDWGDATTTFPRDRWNEPREAL--ASGWDYTFPNGGPRICLGGQFALTASVTIVICQ	480
QY	513	FYSFRLLEGHPVQYRMMTILSLMAHGLKVRV	542
Db	481	ERS-RIEVLHP-----DVITSKNSMKQRM	503
RESULT 6			
CP5L_DEBHA STANDARD; PRT; 519 AA.			
AC	Q9Y757	2001 (Rel. 40, Created)	
DT	16-OCT-2001	(Rel. 40, last sequence update)	
DE	15-MAR-2004	(Rel. 43, last annotation update)	
DE	Cytochrome P450 52A12	(EC 1.14.14.-) (Alkane hydroxylase 1) (Alkane-	
DE	inducible p450alk 1) (DH-ALK2).		
GN	CYP52A12 OR ALK1.		
OS	Debaryomyces hansenii (Yeast) (Torulaspora hansenii).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Debaryomycetes.		
OX	NCBI_TaxId=4959;		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=ATCC 20317;		
RC	MEDLINE=99132287; PubMed=9931473;		
RX	Yadav J.S., Loper J.C.;		
RA	"Multiple p450alk (cytochrome P450 alkane hydroxylase) genes from the		
RT	halotolerant yeast Debaryomyces hansenii.";		
RL	Gene 226:139-146(1999).		
CC	!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system		
CC	catalyzes the terminal hydroxylation as the first step in the		
CC	assimilation of alkanes and fatty acids.		
CC	!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal		
CC	position.		
CC	!- INDUCTION: By N-alkanes.		
CC	!- SIMILARITY: Belongs to the cytochrome P450 family.		
This SWISS-PROT entry is copyright. It is produced through a collaboration			
between the Swiss Institute of Bioinformatics and the EMBL outstation -			
the European Bioinformatics Institute. There are no restrictions on its			
use by non-profit institutions as long as its content is in no way			
modified and this statement is not removed. Usage by and for commercial			
entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; AF103948; AAD22536.1; -.		
DR	HSSP; P14779; 1JJPZ.		
DR	InterPro; IPR001128; Cytochrome_P450.		
DR	Pfam; PF00067; p450; 1.		
DR	PRINTS; PR00385; P450.		
DR	PROSITE; PS00086; CYTOCHROME P450; 1.		
KW	Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.		
FT	METAL. 467 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).		
SQ	SEQUENCE 519 AA; 59218 MW; C56740984B33B0E4 CRC64;		
Query Match 13.7%; Score 423; DB 1; Length 519;			
Best Local Similarity 27.0%; Pred. No. 5.2e-20;			
Matches 154; Conservative 91; Mismatches 210; Indels 116; Gaps 21;			
QY	26	YIALLVVL-----SWILVQWLSLRKQKSPSPV-----IGATVEQLRNYHRHMDLV	74
Db	14	WYGILLAVLIGVHPDYIRIQ--IVMKKLGCVPPEVDEGFFGPKLYTSLKHKEGTLV	71
QY	75	GVLSHRIVTMDPTSYTIADPV-----NVEHLVKTFTNYPKGVVRSYMDVLIG	127
Db	72	NFI-KERTETVGKOTFSFRIAGTPVISTKNPENIKALLATQSFALGTRHAQFPE-LLG	129
QY	128	DGIFNADGELNKRQKTASFEPASKNLDRFSIVFREYSLKLSGLSQAQKAGKVVDMQE	187
Db	130	DGIFLDSGWSKSRAMLRPQFAREQVAHVKSL---EPHIQMLAKHVRRAKGG-AFDVQS	185
QY	189	LYMWTLSICIKGVGVEIGTLPDL-----PNSFAQAFDAANIITLRFIDP	236
Db	186	LFRLTVDSATFLFEGESVESLQDESIGNAKDAVDFDGKAGFAEAFNTAQVYLSIRSLAQ	245
QY	237	LWRIKFEFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIWEVRASGQKQKHDILSRFIE	296
Db	246	----KAYFLVNNKREFSSNEK-VHKFADYVVK-----ALNSSPEELEKH-----	285
QY	297	LGEAGDDGGGF-----GDDKSLRDVLNLFVIAGRDATTTLISWETHMAMSHPDVAEK	348
Db	286	----SQDGYIFLYELVKQTRDPHVLRLDQLNILLAGRDTTAGLLSFTFYELARNPQWLK	341
QY	349	LRRELCAFE-----AERAREGVTILVLCGGADADKAFARVAQFAGLLTYDSLGLVLILH	404
Db	342	LKEEI--YEKFGKGDARLEI-----TFESLKKCEYLK	373
QY	405	ACVTETLRLYPVPODPKGILEDDVLDPG-----TKVRAGGMVTYVPYSGMYNWN	456
Db	374	ALLNEVLRLYPSVPQNFVRVAQKDTSLPRGGPNRDQPIFIAGQTVTYTYVYAMHRDEQFY	433
QY	457	GPDAASFRPERWINEGDAFRNASPFKFTAFQAGPRICLKDSAYLQMKMALILFRFYSF	516
Db	434	GKDSVFRPERWPEPE---TRKLWAFLPNGGPRICLGGQFALTASVYIARLAQL--F	488
QY	517	RLLEGHPVQY--RMMTILSLMAHGLKVRVSRSA	545
Db	489	PNLASHDDDEYPPRKASHLTCHQSEVKISLA	519
RESULT 7			
CP5W_CANAP STANDARD; PRT; 519 AA.			
AC	Q12573;		
DT	15-DEC-1998	(Rel. 37, Created)	
DT	15-DEC-1998	(Rel. 37, last sequence update)	
DT	28-FEB-2003	(Rel. 41, last annotation update)	
DE	Cytochrome P450 52E2	(EC 1.14.14.-) (CYPLIIE2).	
GN	CYP52E2.		
OS	Candida apicola (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.		
OX	NCBI_TaxId=29830;		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=IMET 43747;		
RC	MEDLINE=96367597; PubMed=8771711;		
RX	Lettermose K., Schunck W.H., Asperger O.;		
RA	"Cytochromes P450 of the sophorose lipid-producing yeast Candida		
RT	apicola: heterogeneity and polymerase chain reaction-mediated cloning		
RT	of two genes.";		
RL	Yeast 12:565-575(1996).		
CC	!- SIMILARITY: Belongs to the cytochrome P450 family.		
This SWISS-PROT entry is copyright. It is produced through a collaboration			
between the Swiss Institute of Bioinformatics and the EMBL outstation -			
the European Bioinformatics Institute. There are no restrictions on its			
use by non-profit institutions as long as its content is in no way			
modified and this statement is not removed. Usage by and for commercial			
entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[illegible]

RESULT 8		
ID	CP5C CANMA STANDARD;	PRT; 523 AA.
AC	P16496; P20017;	
DT	01-AUG-1990 (Rel. 15, Created)	
DT	01-AUG-1990 (Rel. 15, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Cytochrome P450 52A3 (EC 1.14.14.-) (CYP11A3) [Alkane-inducible P450-	
DE	ALK1-A] (P450-CM1) (CYP52A3-A) (Cytochrome P-450ALK).	
GN	CYP52A3.	
OS	Candida maltosa (Yeast).	
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.	
OX	NCBI Taxid=5479;	

[1] SEQUENCE FROM N.A.  
RP MEDLINE=92109967; PubMed=1368716;  
RX Ohkuma M., Hakiiji T., Tanimoto T., Schunck W.-H., Mueller H.G.,  
RA Yano K., Takagi M.;  
RR "Evidence that more than one gene encodes n-alkane-inducible  
RT cytochrome P-450s in *Candida maltosa*, found by two-step gene  
RT disruption.";  
RT Agric. Biol. Chem. 55:1757-1764(1991).  
RL [2]  
RN  
RN SEQUENCE FROM N.A.  
RP STRAIN=IAM 12247;  
RC Takagi M., Ohkuma M., Kobayashi N., Watanabe M., Yano K.;  
RA "Purification of cytochrome P-450alk from n-alkane-grown cells of  
RT *Candida maltosa*, and cloning and nucleotide sequencing of the  
RT encoding gene";  
RT Agric. Biol. Chem. 53:2217-2226(1989).  
RL [3]  
RN  
RN SEQUENCE FROM N.A.  
RP STRAIN=EH15D;  
RC MEDLINE=89286595; PubMed=2735924;  
RX Schunck W.-H., Kaergel E., Gross B., Wiedmann B., Mauersberger S.,  
RA Koepke K., Kieselung U., Strauss M., Gaestel M., Mueller H.-G.;  
RR "Molecular cloning and characterization of the primary structure of  
RT the alkane hydroxylating cytochrome P-450 from the yeast *Candida*  
RT *maltosa*.";  
RT Biochem. Biophys. Res. Commun. 161:843-850(1989).  
RL [4]  
RN  
RN SEQUENCE FROM N.A.  
RP MEDLINE=92037671; PubMed=1935956;  
RX Schunck W.-H., Vogel F., Gross B., Kaergel E., Mauersberger S.,  
RA Koepke K., Gengnagel C., Mueller H.-G.;  
RR "Comparison of two cytochromes P-450 from *Candida maltosa*: primary  
RT structures, substrate specificities and effects of their expression  
RT in *Saccharomyces cerevisiae* on the proliferation of the endoplasmic  
RT reticulum.";  
RT Eur. J. Cell Biol. 55:336-345(1991).  
RL [5]  
RN  
RN CHARACTERIZATION.  
RP MEDLINE=96311366; PubMed=8713123;  
RX Zimmer T., Ohkuma M., Ohta A., Takagi M., Schunck W.H.;  
RA "The CYP52 multigene family of *Candida maltosa* encodes functionally  
RT diverse n-alkane-inducible cytochromes P450.";  
RT Biochem. Biophys. Res. Commun. 224:784-789(1996).  
RC [-] FUNCTION: Together with an NADPH cytochrome P450 the enzyme system  
CC catalyzes the terminal hydroxylation as the first step in the  
CC assimilation of alkanes and fatty acids.  
CC [-] CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal  
CC position.  
CC [-] INDUCTION: By alkanes.  
CC [-] SIMILARITY: Belongs to the cytochrome P450 family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D00481; BAA00371.1; -;  
CC EMBL; X51931; CAA36197.1; -;  
CC PIR; B56578; O4CKA3.  
CC HSSP; P14779; LJPZ.  
CC InterPro; IPR001128; Cytochrome\_P450.  
CC Pfam; PF00067; P450; 1.  
CC PRINTS; PR00385; P450.  
CC PROSITE; PS00086; CYTOCHROME P450; 1.  
KW Electron transport; Oxidoreductase; Monooxygenase; Heme;  
KW Transmembrane.  
FT Transmem 17 34 POTENTIAL.  
FT METAL 471 471 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
FT CONFLICT 221 221 L -> I (IN REF. 3).  
FT

CC	CONFLICT	52	52	E -> Q (IN REF. 3).		
CC	CONFLICT	147	147	L -> H (IN REF. 3).		
CC	SEQUENCE	523	AA; 59839	MW; 5B5E7EB432DEE30F	CRC64;	
CC	Query Match					
CC	Best Local Similarity 13.2%; Score 406.5; DB 1; Length 523;					
CC	Matches 145; Conservative 71; Mismatches 187; Indels 113; Gaps 22;					
QY	75	GVLRSRRTVTVDM--PFTSYTIA-----DPNVVHVLKNTNPKGIVYASYMD	123			
DB	72	GQLADYDTITFDKYPNLSSVMTVAGVLKIVFTDENIKAVLATQFNDFAFG-ARAHFD	130			
QY	124	VLLGDGIFNADGELWKRQKTAFFASKNLRFSAIVFREYSKLGLSLOASK--AGK	181			
DB	131	PLLGDDGIFTLDGEGWKLKSRAMLRPQFAREQIAHVKAAL---EPHVQ---ILAKQIKLNGKG	184			
QY	182	VDMOELYMWTLDISICKVGFVEIGTL-----SP-DLP-ENSPAQAFDAANIITL	231			
DB	185	TFDLQELFFRTVDTEFLFGSVHSLYDEKLGIPAPNDIPGRENFAEAFNTSQHYLAT	244			
QY	232	RFIDPL--W--RIKRFPHVGSALLAQSIKLVDEFYTVIRRKAEIVFVRAGKQKMK	287			
DB	245	RTYSQIFMYLTNPKFPRDCN-----AKVHKLQVFNVTALNATEKEY-----BEKSK	291			
QY	288	HDILSRFELGEAGDDGGF-----GDKSLSDVVLNFVIAGRDITATTLTSLWFTH	337			
DB	292	-----GGVFLYELVKQTRDPKVLQDQLINWAGRDITAGLLSPAMF	334			
QY	338	MAMSHPDVAEKLRELCAFEARARESGVTILVLCGGADADDKAPARVAQFAGILLTYDSL	397			
DB	335	ELARNPKIWNKLRV-----EVNFGLEARVDE-----ISFETL	370			
QY	398	GKLVYLHACVTETLRLYPVAPQDPKGILEDDVLP-----DGTK---VRAGMVTYVPSM	449			
DB	371	KCKBYLKAIVNETLRMYPSPVPIERTATRTDTLPRGGGKDGNSPIFVFKGSSVVYVYKT	430			
QY	450	GRMEYNGPDAASPRPERWINEGAFRNASPKETAQAGPRICLIGKDSAYLQMKMALI	509			
DB	431	HLKQFYGEDAYEPFRPERWFPS-----TRKLGWAVLPFNGGPRICLIGQOAFALTEASYVIAR	487			
QY	510	LPRFYSRLLEGHPVQY--RMMTILSMAHGLKVRVS	543			
DB	488	LAQM--FEHLESKDETYPNPKIHLTWNHNEGVIIS	521			
CC	RESULT 9					
ID	CP5F	CANTR	STANDARD;	PRT;	524	AA.
AC	P30608,					
DT	01-APR-1993	(Rel. 25, Created)				
DT	01-APR-1993	(Rel. 25, last sequence update)				
DT	28-FEB-2003	(Rel. 41, last annotation update)				
DE	Cytochrome P450 52A6	(EC 1.14.14.-) (CYPLI1A6)	(Alkane-inducible P450-ALK3).			
GN	CYP52A6.					
OS	Candida tropicalis (Yeast).					
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;					
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.					
OX	NCBI TaxID=5482;					
RN	[1] "					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=ATCC 750;					
RX	MEDLINE=93090274; PubMed=1457045;					
RA	Seghezzi W., Meili C., Ruffiner R., Kuenzi R., Sanglard D.,					
RA	Fiechter A.;					
RT	"Identification and characterization of additional members of the					
RL	cytochrome P450 multigene family CYP52 of Candida tropicalis.";					
CC	DNA Cell Biol. 11:767-780(1992).					
CC	FUNCTION: Together with an NADPH cytochrome P450 the enzyme system					
CC	catalyzes the terminal hydroxylation as the first step in the					
CC	assimilation of alkanes and fatty acids. Preferentially					
CC	hydroxylates hexadecane.					
CC	-!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal					

CC	position.					
CC	-!- INDUCTION: By various alkanes.					
CC	-!- SIMILARITY: Belongs to the cytochrome P450 family.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; Z13010; CAA78354.1; --					
DR	PIR; S22972; S22972.					
DR	HSSP; P14779; LJP2.					
DR	InterPro; IPR001128; Cytochrome_P450.					
DR	PIfam; PF00067; P450; 2.					
DR	PRINTS; PR00385; P450.					
DR	PROSITE; PS00086; CYTOCHROME_P450; 1.					
KW	Electron transport; Oxidoreductase; Monooxygenase; Heme;					
KW	Transmembrane.					
FT	TRANSMEM 17 34 POTENTIAL.					
FT	METAL 472 472 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).					
SQ	SEQUENCE 524 AA; 59927 MW; 31283F5B2135749 CRC64;					
	Query Match 13.1%; Score 402; DB 1; Length 524;					
	Best Local Similarity 27.0%; Pred. No. 1.2e-18;					
	Matches 137; Conservative 70; Mismatches 181; Indels 120; Gaps 20;					
QY	46	KQKG-----PRSNPVIGATVEOLNRYHRMDWLVGYLSRHRVTIVDMFTSYTIAFPVNV	101			
DB	67	KVKGNGRLAREFWPI--KTFFDYPN-HTFYMSIIGALKXIVLTVIQE-----NI	110			
QY	102	EHVLKTNFTNYPKGIYVRSYMDVLGDIENAGELWKRKORTKATFASFASKNLRFDSAIV	161			
DB	111	KAVLATQFTDSLGRHAHYF--LLGDIIFTLDEGKHSRAMLRPQFARDQIGHVKAL--	168			
QY	162	FREYSKLGLSLOASK--AGKVYVDMOELYNMRLTDSICKVGFVEIGTL-----SP	211			
DB	169	--EPHIQ---ILAKQIKLNGKKTFDIQELPFRFTVDATFELFGESVHSLYDEKLGITP	223			
QY	212	-DLP-ENSPAQAFDAANIIITLRFIDPLWLRKRFHHVGSALLAQSIK---LVDEFTYSV	266			
DB	224	NEIPGRDNFATFNTSQHYLATRYTSQTF-----YFLTNPKFRDCNAKVHLYAKYFNKA	279			
QY	267	IRRRKAEIVFVRAG-----KQEKMKHDIILSRFELGEAGDDGGFGDKSLRDVWL	318			
DB	280	LNFTPEELIEKSKGYVELYELVKQTR-----DPKVLQDQLL	316			
QY	319	NFVIAGRDITATTLTSLWFTHMAMSHPDVAEKLREL---CAFEARAREEGVTULVLCGA	374			
DB	317	NIMWAGRDITAGLLSFAMFELARHPFELWKLREIEVNFVGVGEESRVEE-----	365			
QY	375	DADDKAFARVAQFAGLLTYDSLGLVYLHACVTETLRLYPVAPQDPKGILLEDDVL---	430			
DB	366	-----ITFESLKECYLKALINLETLMYPSVNVNSTATRTDTLTPRG	408			
QY	431	-PDGTK---VRAGMVTYVPSYMGMEYNWGPDAASPRPERWINEGAFRNASPKFETAF	486			
DB	409	GGNGTDPIFIPKGSVAVIVYKTHREBYGKADDFRPERWFPS-----TKKLGWAVYVF	465			
QY	487	QAGPRICLIGKDSAYLQMKMALIALFRFY	514			
DB	466	NGGPRICLIGQOQFALTEASYVITRLVQMF	493			
CC	RESULT 10					
ID	CP5E	CANNA	STANDARD;	PRT;	523	AA.
AC	P24458;					
DT	01-MAR-1992	(Rel. 21, Created)				
DT	01-MAR-1992	(Rel. 21, last sequence update)				
DT	28-FEB-2003	(Rel. 41, last annotation update)				



Cytochrome P450 52A5 (EC 1.14.14.-) (CYPLI1A5) (Alkane-inducible P450-  
 DE ALK2-A) (CYP52A3-B).  
 GN CYP52A5.  
 OS Candida maltosa (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5479;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9210967; PubMed=1368716;  
 RA Okuma M., Hiki T., Tanimoto T., Schunck W.H., Mueller H.G.,  
 RA Yano K., Takagi M.;  
 RT "Evidence that more than one gene encodes n-alkane-inducible  
 RT cytochrome P-450s in Candida maltosa, found by two-step gene  
 RT disruption.";  
 RL Agric. Biol. Chem. 55:1757-1764 (1991).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96311366; PubMed=8713123;  
 RA Zimmer T., Okuma M., Ohta A., Takagi M., Schunck W.H.;  
 RA "The CYP52 multigene family of Candida maltosa encodes functionally  
 RT diverse n-alkane-inducible cytochromes P450.";  
 RL Biochem. Biophys. Res. Commun. 224:784-789 (1996).  
 CC -!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system  
 CC catalyzes the terminal hydroxylation as the first step in the  
 CC assimilation of alkanes and fatty acids.  
 CC -!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal  
 CC position.  
 CC -!- INDUCTION: By alkanes.  
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: D12475; BAA02041.1; -;  
 EMBL: S77461; AAC60531.1; -;  
 PIR: JQ1039; JQ1039.  
 HSP: P44779; I3P2.  
 InterPro: IPR001128; Cytochrome\_P450.  
 Pfam: PF00067; P450; 1.  
 PRINTS: PR00385; P450.  
 PROSITE: PS00086; CYTOCHROME P450; 1.  
 Electron transport; Oxidoreductase; Monooxygenase; Heme;  
 Transmembrane.  
 TRANSMEM 17 34 POTENTIAL.  
 FT METAL 471 471 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 523 AA; 59868 MW; DB357103097F7FE8 CRC64;  
 Query Match 13.0%; Score 400.5; DB 1; Length 523;  
 Best Local Similarity 27.0%; Pred. No. 1.4e-18;  
 Matches 138; Conservative 71; Mismatches 193; Indels 109; Gaps 18;  
 66 YHRMDLWGLVGRHRTYVDMPFTSYTYIADPNVVEHLKTNFTNPKGIVRSYMDVL 125  
 87 HHSFYMTVAGMLKIVLTV-----DPENIKAVLATQFNDPALG-ARHAHFDPL 132  
 126 LGDGFNADGLWRKORTASFPASKNLDKDFSAIVREYSKLGLSLQASK--AGKV 183  
 133 LGDGIPTLDGEGWKHSRAMLRLPQAPQECIAHVKAL---EPHVQ---VLAKQIKLNGKETF 166  
 184 DMQELYMRWTLDSCKVGFGEIGTLSPDL-----PEN-----SFAQAFDAANIIILRF 233  
 187 DLQELFRFTVDTEFLFGESVHSLDDEKLGVPFPNNIPGRENFAKAFNTSQHLYATRT 246  
 234 IDPLMRKRPFPVHGEALL---AQSIKIVDEFTYSVIRRKAEIVFVRASG-----K 282  
 247 YSQMF---YFLTPKPEFRDNCNAKVHKLQAYFVNKALDASDEVAEKSKGVPFLVLVK 302

283 QEKXKHDILSRFTELGEAGDGGGDDKSLRDVNLVNFVIGRDTTATITLSWFTMAMSH 342  
 303 QTR-----DPKVLQQLNIMVAGRDITAGLLSFAFMELARN 339  
 343 PDVAEKLRELCAFEAREAREEGVTVLVCGGADADDKFAARVAQFAGLLTYSGLKLVY 402  
 340 PKIWNKLRBI-----EVNFGLGGEARVDE-----ISFETLKCEY 375  
 403 LHACVITETIRLYPAVPODPKGLLEDDVLP-----DGTK-----VRAGGMVTVPVSMGRMBY 454  
 376 LKAVINETLRMPSPVYNFTATRTDTLPRGGKDGSPFVFKGSSVTVTVYKTHLEE 435  
 455 NWGPDAAASFRPERINEDGAFRNASPEKFTAFQAGPRICLGKDSAYLQMKMALAILFRFY 514  
 436 YYGKDAYEFRPERWFPS---TRKLGWAYVFPENGPRICLGQGFALTEASYVITRLAQM- 491  
 515 SFRLEGEHPVOY--RMVTILSMAHGLKVRVS 543  
 492 -FEHLESKDQETYPNPKCHILTMHNEGVS 521  
 RESULT 11  
 C4E1 DROME  
 ID C4E1 DROME STANDARD; PRT; 531 AA.  
 AC Q9V4T5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable cytochrome P450 4e1 (EC 1.14.-.-) (CYP1E1).  
 GN CYP4E1 OR CG2062.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan X.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agyarani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein E., Center A., Chandra I.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Caudy J., Davies P.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
CC -!- FUNCTION: May be involved in the metabolism of insect hormones and  
CC in the breakdown of synthetic insecticides (By similarity).  
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
CC oxidized flavoprotein + H(2)O.  
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum  
CC (potential).  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: A5003837; AAF59090.1; --  
CC FlyBase: FBgn0015034; Cyd4el.  
CC InterPro: IPR001128; Cytochrome\_P450.  
CC Pfam: PF00067; P450; 1.  
CC PRINTS: PR00385; CYTOCHROME\_P450.  
CC PROSITE: PS00086; CYTOCHROME\_P450; 1.  
CC Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;  
KW Endoplasmic reticulum; Hypothetical protein.  
KM 444 444 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
FT METAL 444 444 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 531 AA; 61086 MW; 42D45A86B5DB905 CRC64;  
  
Query Match 12.9%; Score 398.5; DB 1; Length 531;  
Best Local Similarity 25.2%; Pred. No. 2e-18;  
Matches 148; Conservative 91; Mismatches 217; Indels 131; Gaps 24;  
  
QY 26 YTALLVLVLSW-----ILVQWSLRKQKGRSFGVIGATVEQLRN-----YHRHWD----- 72  
DB 8 FLALPLFLVTFELGLLRKRLMKLPQGPSMLPLVG-NAHQMGNTPEILNFFGWNHWEY 66  
QY 73 -----LVGYLSRHRVTVDMPFTSYTVIADPVNVEHLKTNFTNPKGIVYRSYMDVL 125  
DB 67 GKDNFRYWTGYISN-----IMVTNPKYMEFILSSQ-TLISKSDVY-DLTHWN 111  
QY 126 LGDGFENADGELWRKQRTASFEFASKNLDRDFAIVFREYSKLGLSKGLSQASKAGKVWDM 185  
DB 112 LGLGLLTSTGSKWKKRKMTPAFHENILQDPHE-VNENSTKFIQDLKKVADGGNIQDF 170  
QY 186 QELYMRMTLSDICKVGFVIGTGLSPDLSPNSAQAFDAANIIITLRFIDPLWRKRFH 245  
DB 171 QEEAHVLTLDVICDTANGVSNAM--ENRSSVVQAFKIDITYIKRASFSPWKRNYLFH 228  
QY 246 VGSE-ALLAQSIKLVDEFTYSVIRRRKAEIVFVRASGQKMKHDLISR-----FIELGEA 300  
DB 229 FAPFPEYSKTLTQLDFTNEIIAKR-----IEVRKSGLEVGIKADFSRKKMAFLDTLS 284  
QY 301 GDGGGFGDDKSLRDVVLNFIAGRTDTATTLTGWFTHMVSHPDVAFKRLREICAFEAER 360  
DB 285 SKVDGRPLTSQELVEEVSTFMFEGHDTTSGVGFVAVLLSRHPDEQEKLFNEQC-----DV 340  
QY 361 AREGVTILVCGADADDKAPARVAQFAGLLIYDSLGKLVYLHACVTEITRLYPVP-- 418  
DB 341 MGASGL-----GRDA-----TFQEIISTMKHLDLFTKEAQRLYSPVFFI 378  
QY 419 ---QDPKGILEDDVLPGTKVRAGMVTYVPYSNGRMEYN--WGPDAASFRPERWINEDG 473  
DB 379 GRFTEKDYVIGDILVPKGTINLGLIM-----LGYNDRVPKDPKHPQPERFDE-- 427  
QY 474 AFRNASPFKFTAFQAGPRICLGKDSAYLOKMAJAILFRFYSFRL----- 519  
DB 428 ---KPGFPEYVFPFAGPRNCIGQKFALEIKTVVSKLIIR--NFEVLPAIDELVSKOGYIS 482  
QY 520 -----EGH-----PVQYRMWTLISMAHGLKRVV-SRAVC 547

DB 483 TTLGLQPAEKSRDAHNHXYDILSASMTLKS-ENGLHLRMKQRLVC 528  
  
RESULT 12  
ID CP5M DEBHA STANDARD; PRT; 519 AA.  
AC Q9Y758;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Cytochrome P450 52A13 (BC 1.14.14.-) (Alkane hydroxylase 2) (Alkane-  
DE inducible p450alk 2) (DH-ALK2).  
GN CYP52A13 OR ALK2.  
GN Debaryomyces hansenii (Yeast) (Torulaspora hansenii).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
OX NCBI\_TaxID=4959;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 20317;  
RC MEDLINE=99132287; PubMed=9931473;  
RA Yadav J.S., Loper J.C.;  
RT "Multiple p450alk (cytochrome P450 alkane hydroxylase) genes from the  
RT halotolerant yeast *Debaryomyces hansenii*.";  
RL Gene 226:139-146(1999).  
CC -!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system  
CC catalyzes the terminal hydroxylation as the first step in the  
CC assimilation of alkanes and fatty acids.  
CC -!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal  
CC position.  
CC -!- INDUCTION: By N-alkanes.  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF103949; RAD22537.1; --  
CC HSSP: P14779; IJPZ.  
CC InterPro: IPR001128; Cytochrome\_P450.  
CC Pfam: PF00067; P450; 1.  
CC PRINTS: PR00385; P450.  
CC PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.  
FT METAL 466 466 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 519 AA; 59607 MW; 1CE7D1EE2CE5B5A CRC64;  
  
Query Match 12.8%; Score 395; DB 1; Length 519;  
Best Local Similarity 25.1%; Pred. No. 3.2e-18;  
Matches 141; Conservative 96; Mismatches 222; Indels 102; Gaps 17;  
  
QY 26 YTALLVLVLSWILVQWSLRKQKGR-----SNPVGATVEQLRNHRMHD 71  
DB 16 YVVISALLVSFFIAHKISVARFKATHNCAASPEYKYNWFSPLLYRLI-QVREGRLLD 74  
QY 72 WLVGYSLRHRTVTDMPFTSYTVI--ADPVNVEHLKTNFTNPKGIVYRSYMDVLGDS 129  
DB 75 FQKIYDDVKALTFVIKVGVPVITRDENKKAVALQTFNDPALG-TRAHFKPLGDS 133  
QY 130 IFNADGELWRKQRTASFEFASKNLDRDFAIVFREYSKLGLSKGLSQAASKAGKVWDMQELY 189  
DB 134 IFPLDGNKQKSRMLRPQFSREQVAHQAL--EPHLQELAKHRLAD--GETINIQDLF 189  
QY 190 MRMTLSDICKVGFVIGTGLSP-----DLP-----ENSPAQAFDAANIIITLRFIDPLMR 239  
DB 190 FKLTVDVTFATEFLFGQSVSLKDAAINDPPTEDPDGRSSFANSFNTACTYLGTRAY----- 244  
QY 240 IKKFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIVFVRASGQKMKHDLISRFIELGE 299

Matches 131; Conservative 74; Mismatches 183; Indels 88; Gaps 17;

QY 97 DPVNEVHLKTNFTNPKGIVVRSYMDVLGDIQFNADGELWKKQRTAFSEFASKWLRLD 156  
 Db 101 DPNIKAMLATQENDFTLG-QRLSYFAPLIGKIGFTLDGEGMKHSLRMLRPPQSRDOVGH 159  
 QY 157 FSAI-----VFRYSYSLKLSGILSOAKGKVVDMQELYMWTLDLSICKVGFVGEIGTSLP 211  
 Db 160 VKMLEPHFOLLKXHIK-----NKGFFDIQELFFFTVDSATEFLFGSVSLKD 210  
 QY 212 D-----LPENSAFAQAFDAAMIIITLRFIDPLWRIKFRPHVSEALLAQSIKLVLD 260  
 Db 211 ESTGYDOEBIDFAGKDFAEAFKNSQVYLSTRSL-----LQLLYLWLVNSDFKRCNKIVH 265  
 QY 261 EFTYSVIRRKAEIVVRSAGKQKQKHDLISFIELGEAGDDGGGDDGDKSLRDVVLNF 320  
 Db 266 KFSYVIKK-----ALTATPELEKH-----SSYIFLYELAKQ---TRDPVILRDQSLNI 312  
 QY 321 VIAGDRTTATLSWFTWAMSHDPVAKLRRLCAFEARERBEGVTLVLCG---GADAD 377  
 Db 313 LLAGDRTTAGLLSPAVFELGRNPEVWSKLREI-----GDKFGLDPD 354  
 QY 378 DKAFARVAQFAGLLTYDSLGLKLYLHACVETETLRLYPVQDPKGLIEDDVLDPG--- 433  
 Db 355 -----SRIED-----ISFELLKLCYLVKAVINETLRLYPSVPRNGRFAAANTLPHGGGPD 405  
 QY 434 -----TKVRAGMVTYVYPSYSGRMVYNGPDAASFRPERKINEDGAFNAPSPFKETAQAG 489  
 Db 406 GMSFILVRKGTVMYSYVALQDRDEKYGKADNEFRPERWFEPE---VRKLG-WAFLPFGG 462  
 QY 490 PRICLGKDSAYLQMKALAILFR-FYSFRLLEGHFVQVYMMTILSMA--HGLKVRV 542  
 Db 463 PRICLQGFALTEASYVLVRLIQSFTEILESPDAPYPPAKLTLTMCILFDGAPVRI 518

RESULT 14

CP5K\_CANMA STANDARD; PRT; 519 AA.

AC Q12589; (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE "CYP52 (cytochrome P450alk) multigene family in Candida maltosa: identification and characterization of eight members."  
 DE P450-ALK8).  
 GN CYP52A11.  
 OS Candida maltosa (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5479;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95162975; PubMed=7865134;  
 RA Okuma M., Muraoka S., Tanimoto T., Fujii M., Ohta A., Takagi M.;  
 RT "CYP52 (cytochrome P450alk) multigene family in Candida maltosa: identification and characterization of eight members."  
 RL DNA Cell Biol. 14:163-173(1995).  
 CC -!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system catalyzes the terminal hydroxylation as the first step in the assimilation of alkanes and fatty acids.  
 CC -!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal position.  
 CC -!- INDUCTION: By N-alkanes.  
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D12719; BAA02214.1; -;  
 CC PIR: JS0726; JS0726.  
 CC HSP: P14779; LJPZ.  
 CC InterPro: IPR001128; Cytochrome\_P450.  
 CC Pfam: PF00067; P450; 1.  
 CC PRINTS: PR00385; P450.  
 CC PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.  
 FT METAL 466 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 519 AA; 59476 MW; 70A9430ADEE38096 CRC64;

Query Match 12.6%; Score 387; DB 1; Length 519;  
 Best Local Similarity 27.5%; Pred. No. 1.1e-17;

Matches 131; Conservative 74; Mismatches 183; Indels 88; Gaps 17;

QY 97 DPVNEVHLKTNFTNPKGIVVRSYMDVLGDIQFNADGELWKKQRTAFSEFASKWLRLD 156  
 Db 101 DPNIKAMLATQENDFTLG-QRLSYFAPLIGKIGFTLDGEGMKHSLRMLRPPQSRDOVGH 159  
 QY 157 FSAI-----VFRYSYSLKLSGILSOAKGKVVDMQELYMWTLDLSICKVGFVGEIGTSLP 211  
 Db 160 VKMLEPHFOLLKXHIK-----NKGFFDIQELFFFTVDSATEFLFGSVSLKD 210  
 QY 212 D-----LPENSAFAQAFDAAMIIITLRFIDPLWRIKFRPHVSEALLAQSIKLVLD 260  
 Db 211 ESTGYDOEBIDFAGKDFAEAFKNSQVYLSTRSL-----LQLLYLWLVNSDFKRCNKIVH 265  
 QY 261 EFTYSVIRRKAEIVVRSAGKQKQKHDLISFIELGEAGDDGGGDDGDKSLRDVVLNF 320  
 Db 266 KFSYVIKK-----ALTATPELEKH-----SSYIFLYELAKQ---TRDPVILRDQSLNI 312  
 QY 321 VIAGDRTTATLSWFTWAMSHDPVAKLRRLCAFEARERBEGVTLVLCG---GADAD 377  
 Db 313 LLAGDRTTAGLLSPAVFELGRNPEVWSKLREI-----GDKFGLDPD 354  
 QY 378 DKAFARVAQFAGLLTYDSLGLKLYLHACVETETLRLYPVQDPKGLIEDDVLDPG--- 433  
 Db 355 -----SRIED-----ISFELLKLCYLVKAVINETLRLYPSVPRNGRFAAANTLPHGGGPD 405  
 QY 434 -----TKVRAGMVTYVYPSYSGRMVYNGPDAASFRPERKINEDGAFNAPSPFKETAQAG 489  
 Db 406 GMSFILVRKGTVMYSYVALQDRDEKYGKADNEFRPERWFEPE---VRKLG-WAFLPFGG 462  
 QY 490 PRICLGKDSAYLQMKALAILFR-FYSFRLLEGHFVQVYMMTILSMA--HGLKVRV 542  
 Db 463 PRICLQGFALTEASYVLVRLIQSFTEILESPDAPYPPAKLTLTMCILFDGAPVRI 518

RESULT 14

CPFC\_HUMAN STANDARD; PRT; 524 AA.

AC Q9HCS2; Q9HCS1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 4F12 (EC 1.14.14.1) (CYP4F12).  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND CHARACTERIZATION.  
 RC TISSUE=Liver;  
 RX MEDLINE=21092895; PubMed=11162607;  
 RA Bylund J., Bylund M., Olin E.H.;  
 RT "cDNA cloning and expression of CYP4F12, a novel human cytochrome P450."  
 RL Biochem. Biophys. Res. Commun. 280:892-897(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, CHARACTERIZATION, AND VARIANT LEU-13.  
 RC TISSUE=Small intestine;  
 RX MEDLINE=21092733; PubMed=11162645;  
 RA Hashizume T., Imaoka S., Hiroi T., Terauchi Y., Fujii T., Miyazaki H.;  
 RT "cDNA cloning and expression of a novel cytochrome p450 (cyp4f12) from human small intestine."  
 CC -!- FUNCTION: Catalyzes leukotriene B(4) omega-hydroxylation and arachidonic acid omega-hydroxylation but with an activity much lower than that of CYP4F2. Catalyzes the hydroxylation of the antihistamine ebastine.  
 CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH + oxidized flavoprotein + H(2)O.  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By

```
CC CC similarity).
CC CC -!- TISSUE SPECIFICITY: Expressed in small intestine, liver, colon and
CC CC heart.
CC CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; AY008841; AAG33247.1; -
CC CC EMBL; AB035130; BAB18269.1; -
CC CC EMBL; AB035131; BAB18270.1; -
CC CC PIR; JC7594; JC7594.
CC CC PIR; JC7598; JC7598.
CC CC Genew; HGNC:18957; CYP4F12.
CC CC InterPro; IPR001128; Cytochrome_P450.
CC CC PRINTS; PR00385; P450.
CC CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum; Polymorphism.
FT DOMAIN 442 447 MEMBRANE (RETICULUM ENDOPLASMIC) -BOUND
FT METAL 468 468 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT VARIANT 13 13 P -> L.
FT VARIANT 76 76 /FTID=VAR_013244.
FT VARIANT 90 90 D -> N (in dbSNP:609636).
FT VARIANT 90 90 V -> I (in dbSNP:609290).
FT VARIANT 188 188 /FTID=VAR_013246.
FT FT R -> C (in dbSNP:2285888).
FT FT /FTID=VAR_013247.
FT FT /FTID=VAR_013247.
FT FT A71335B02E367C0C CRC64;
SQ SEQUENCE 524 AA; 60309 MW; 7A1335B02E367C0C CRC64;
Query Match 12.6%; Score 386.5; DB 1; Length 524;
Best Local Similarity 24.7%; Pred. No. 1.1e-17;
Matches 144; Conservative 99; Mismatches 234; Indels 107; Gaps 19;
QY 9 LTPATSPFFFLAGPHKXIALLVLSWLV-----QRSLKXKQKPSW-- 53
DB 11 LRPVAMSP-----WLLLLVVGSWLLARILAWTYAFYNNCRRLQCFQPPKRNFW 61
QY 54 ---PVIGATVEQLRNYHRMDWLVGYLSRHTVTVDMPFTSYTIAQPVNVVHVLKTNFT 110
DB 62 GHGLITPTBEGKLDSTQMS-----ATYSQFTTWLG-PFIPFVILCHPDTRISITNSAA 116
QY 111 NYPKGIYVRSYMDVLLGDGFNADGELWRKORKTASFEFAKSLNRDFAIVFREYSKLS 170
DB 117 IAPKDNLFIRELKPWLGEGILLSGDKWKRHRRLTTPAFHFNILKSYTIFPKNSANIMLD 176
QY 171 GILSOAKAGKVMQELMYRMWTLDSICKVGFVEIGTSLDLPENSEAQAFAADANII-- 228
DB 177 KWQHLAGSGSSRLDMFHEHISLMTLDSLOKCTFSTF--SHCQERPEYATITLSELSVEK 234
QY 229 ---ITLRFIDPLWIK---REFHVGSEALLAQSKLVDFEYTSVIRRRKABIVEVRASGK 282
DB 235 RSQHILQHMDELYLISHDGRPH-----RACELVHDFDVAIVRERRRL---PTQGI 283
QY 283 QEKMKHDLISFIEJGE-----AGDDGGFGDGDKSLRDVVLNPFVIAGRTTATTLSWFTHM 338
DB 284 DDFKDKAKSKTLDFIDVLLLSKDEDKGKALSDEDIAREADTFMFGGHDTTASGLSWLVN 343
QY 339 AMSHPDVAEKARELCFAEAREAREEGVTVLCCGADADDKFAARVAQFAGLLTYDSIG 398
DB 344 LARHPYEQRCQEQVEQLLKDRDPKE-----ISWDLA 376
QY 399 KLVYLHACVTEILRYLPAVQDPKGLEDDVLPGDKTKVRAGGMVTYVPSVGMGRMEYN--- 455
DB 377 QLFFLTMCVKESLRHLPAPFISRCTQDVLVLPDGRVIPKG---ITCLIDIIG-VHNTTV 433
```

```
QY 456 WGPDAASFRPERWINEGAFENASPPKFTAFQAGPRICLGDSDAYLQWKMALAILPRFYS 515
DB 434 W-PDPEVYDFRFPDENS--KGRSPLAFIPSGFRNCIGQAFAMKVVLM--LLH 488
QY 516 FRLEGEHFVQVRMTIISMAHGLKVRVSRVCHGDMDIVPLN 559
DB 489 FRFLPDHTPRKLEIMRA-----EGGLMRVVEPLN 520
RESULT 15
CP5D CANNA
ID CP5D CANNA STANDARD; PRT; 538 AA.
AC P16141.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 52A4 (EC 1.14.14.-) (CYPL11A4) (Alkane-inducible P450-
DE ALK3-A) (P450-CM2).
GN CYP52A4.
OS Candida maltosa (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5479;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EH15D;
RX MEDLINE=89286595; PubMed-2735924;
RA Schunck W.-H., Kaergel E., Gross B., Wiedmann B., Mauersberger S.,
RA Koepke K., Kiessling U., Strauss M., Gaestel M., Mueller H.-G.;
RT "Molecular cloning and characterization of the primary structure of
RT the alkane hydroxylating cytochrome P-450 from the yeast Candida
RT maltosa."; Biochem. Biophys. Res. Commun. 161:843-850(1989).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9129697; PubMed=2039569;
RA Okuma M., Tanimoto T., Yano K., Takagi M.;
RT "cyp52 (cytochrome P450alk) multigene family in Candida maltosa:
RT molecular cloning and nucleotide sequence of the two tandemly
RT arranged genes."; DNA Cell Biol. 10:271-282(1991).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96311366; PubMed=8713123;
RA Zimmer T., Okuma M., Ohta A., Takagi M., Schunck W.H.;
RT "The CYP52 multigene family of Candida maltosa encodes functionally
RT diverse n-alkane-inducible cytochromes P450.";
RL Biochem. Biophys. Res. Commun. 224:784-789(1996).
CC -!- FUNCTION: Together with an NADPH cytochrome P450 the enzyme system
CC catalyzes the terminal hydroxylation as the first step in the
CC assimilation of alkanes and fatty acids.
CC -!- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal
CC position.
CC -!- INDUCTION: By alkanes.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51932; CAA36198.1; -
CC PIR; S08668; O4CKA4.
CC HSP; P14779; 1JPZ.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Electron transport; Oxidoreductase; Monooxygenase; Heme;
KW
```

Search completed: March 2, 2004, 15:28:12  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:29:15 ; Search time 38 Seconds  
(without alignments)  
3267.319 Million cell updates/sec

Title: US-10-021-657-2  
Perfect score: 3079  
Sequence: 1 EFGTREAHLTPATSPFFPL.....ICMHACKGRWVSLVAVLKP 588

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3079	100.0	588	13	US-10-021-657-2
2	3079	100.0	588	14	US-10-412-000-2
3	770	25.0	158	13	US-10-021-657-4
4	770	25.0	158	14	US-10-412-000-4
5	401	13.0	492	15	US-10-369-493-3749
6	396	12.9	517	9	US-09-911-781-32
7	396	12.9	517	10	US-09-976-800-100
8	396	12.9	517	14	US-10-138-838-100
9	396	12.9	517	14	US-10-139-031-100
10	396	12.9	517	14	US-10-138-905-100
11	396	12.9	517	14	US-10-138-916-100
12	396	12.9	517	14	US-10-139-296-100
13	396	12.9	517	14	US-10-139-218-100
14	396	12.9	517	14	US-10-400-902-32
15	396	12.9	517	14	US-10-405-660-100

16	396	12.9	517	15	US-10-138-898-100	Sequence 100, App
17	395	12.8	517	10	US-09-976-800-101	Sequence 101, App
18	395	12.8	517	14	US-10-138-838-101	Sequence 101, App
19	395	12.8	517	14	US-10-139-031-101	Sequence 101, App
20	395	12.8	517	14	US-10-138-905-101	Sequence 101, App
21	395	12.8	517	14	US-10-138-916-101	Sequence 101, App
22	395	12.8	517	14	US-10-139-296-101	Sequence 101, App
23	395	12.8	517	14	US-10-139-218-101	Sequence 101, App
24	395	12.8	517	14	US-10-405-660-101	Sequence 101, App
25	395	12.8	517	15	US-10-138-898-101	Sequence 101, App
26	393.5	12.8	546	15	US-10-291-265-405	Sequence 405, App
27	393	12.8	522	10	US-09-976-800-97	Sequence 97, App1
28	393	12.8	522	14	US-10-138-838-97	Sequence 97, App1
29	393	12.8	522	14	US-10-139-031-97	Sequence 97, App1
30	393	12.8	522	14	US-10-138-905-97	Sequence 97, App1
31	393	12.8	522	14	US-10-138-916-97	Sequence 97, App1
32	393	12.8	522	14	US-10-139-296-97	Sequence 97, App1
33	393	12.8	522	14	US-10-139-218-97	Sequence 97, App1
34	393	12.8	522	14	US-10-405-660-97	Sequence 97, App1
35	393	12.8	522	15	US-10-138-898-97	Sequence 112, App
36	389	12.6	517	10	US-09-976-800-112	Sequence 112, App
37	389	12.6	517	14	US-10-138-838-112	Sequence 112, App
38	389	12.6	517	14	US-10-139-031-112	Sequence 112, App
39	389	12.6	517	14	US-10-138-905-112	Sequence 112, App
40	389	12.6	517	14	US-10-138-916-112	Sequence 112, App
41	389	12.6	517	14	US-10-139-296-112	Sequence 112, App
42	389	12.6	517	14	US-10-139-218-112	Sequence 112, App
43	389	12.6	517	14	US-10-405-660-112	Sequence 112, App
44	389	12.6	517	15	US-10-138-898-112	Sequence 112, App
45	389	12.6	523	10	US-09-976-800-95	Sequence 95, App1

ALIGNMENTS

RESULT 1  
US-10-021-657-2  
; Sequence 2, Application US/10021657  
; Publication No. US20020083463A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; TITLE OF INVENTION: METHOD OF USING SAME  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021,657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 588  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-021-657-2

Query Match 100.0%; Score 3079; DB 13; Length 588;  
Best Local Similarity 100.0%; Pred. No. 1.2e-313;  
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	EFGTREAHLTPATSPFFPLAGPHKYIALLLVLSWILVQWLSRKQKQPRSPWPVIGATV	60
Db	1	EFGTREAHLTPATSPFFPLAGPHKYIALLLVLSWILVQWLSRKQKQPRSPWPVIGATV	60
Qy	61	EQLNRYHMDWLVGYLSRHRVTVDMPFTSYTYIADPVNVEHVLKTNFTNPKGIVYRS	120
Db	61	EQLNRYHMDWLVGYLSRHRVTVDMPFTSYTYIADPVNVEHVLKTNFTNPKGIVYRS	120
Qy	121	YMDVLLGDGIFNAGELWVRKQKRTASFEFASKNLRDPSAIVFREYSLSKLSGILSQASKAG	180
Db	121	YMDVLLGDGIFNAGELWVRKQKRTASFEFASKNLRDPSAIVFREYSLSKLSGILSQASKAG	180



QY 181 KVVDMQELYNRMWLDLSICKVGFVGEIGTSLSPDLPENSFAQAFDAANIITLRFIDPLWRI 240  
DB 181 KVVDMQELYNRMWLDLSICKVGFVGEIGTSLSPDLPENSFAQAFDAANIITLRFIDPLWRI 240  
QY 241 KRFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDILSRFIELGEA 300  
DB 241 KRFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDILSRFIELGEA 300  
QY 301 GDDGGGFGDDKSLRDVVLNFVIAGRDATTATLSWFTHMAMSHPDVAEKLRRRELCAFEAER 360  
DB 301 GDDGGGFGDDKSLRDVVLNFVIAGRDATTATLSWFTHMAMSHPDVAEKLRRRELCAFEAER 360  
QY 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTTETLRLYPVPOD 420  
DB 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTTETLRLYPVPOD 420  
QY 421 PKGILEDDVLPDGTGKVRAGGWTVPYSGRMEYNWGPDAASFRPERWINEGAFRNASP 480  
DB 421 PKGILEDDVLPDGTGKVRAGGWTVPYSGRMEYNWGPDAASFRPERWINEGAFRNASP 480  
QY 481 FKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGGHPVQYRMWTTILSMAGHLKV 540  
DB 481 FKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGGHPVQYRMWTTILSMAGHLKV 540  
QY 541 RVSRAVCHGDLMDIVPLNPRQITVLQICMHACKGRWVSLVAVLKP 588  
DB 541 RVSRAVCHGDLMDIVPLNPRQITVLQICMHACKGRWVSLVAVLKP 588

## RESULT 2

US-10-412-000-2  
; Sequence 2, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; PRIOR FILING DATE: 2003-04-11  
; PRIOR FILING DATE: 2003-04-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 588  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-412-000-2

Query Match 100.0%; Score 3079; DB 14; Length 588;  
Best Local Similarity 100.0%; Pred. No. 1-2e-313;  
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EFGTREAHLTPATSPFPPLAGPHKYIALLLVLSWILVQWRSIRKQKGRSPWPVIGATV 60  
DB 1 EFGTREAHLTPATSPFPPLAGPHKYIALLLVLSWILVQWRSIRKQKGRSPWPVIGATV 60  
QY 61 EQLRNYHRMHDWLVGYLSRHTVTVDMPFTSYTIAADPVNVEHLKTNFTNPKGIYVRS 120  
DB 61 EQLRNYHRMHDWLVGYLSRHTVTVDMPFTSYTIAADPVNVEHLKTNFTNPKGIYVRS 120  
QY 121 YMDVLLGDGIFNADGELWRKORTASPEFASKNLRDFAIVFREYSKLGSLSQASKAG 180  
DB 121 YMDVLLGDGIFNADGELWRKORTASPEFASKNLRDFAIVFREYSKLGSLSQASKAG 180  
QY 181 KVVDMQELYNRMWLDLSICKVGFVGEIGTSLSPDLPENSFAQAFDAANIITLRFIDPLWRI 240  
DB 181 KVVDMQELYNRMWLDLSICKVGFVGEIGTSLSPDLPENSFAQAFDAANIITLRFIDPLWRI 240

QY 241 KRFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDILSRFIELGEA 300  
DB 241 KRFFHVGSEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKHDILSRFIELGEA 300  
QY 301 GDDGGGFGDDKSLRDVVLNFVIAGRDATTATLSWFTHMAMSHPDVAEKLRRRELCAFEAER 360  
DB 301 GDDGGGFGDDKSLRDVVLNFVIAGRDATTATLSWFTHMAMSHPDVAEKLRRRELCAFEAER 360  
QY 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTTETLRLYPVPOD 420  
DB 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTTETLRLYPVPOD 420  
QY 421 PKGILEDDVLPDGTGKVRAGGWTVPYSGRMEYNWGPDAASFRPERWINEGAFRNASP 480  
DB 421 PKGILEDDVLPDGTGKVRAGGWTVPYSGRMEYNWGPDAASFRPERWINEGAFRNASP 480  
QY 481 FKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGGHPVQYRMWTTILSMAGHLKV 540  
DB 481 FKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGGHPVQYRMWTTILSMAGHLKV 540  
QY 541 RVSRAVCHGDLMDIVPLNPRQITVLQICMHACKGRWVSLVAVLKP 588  
DB 541 RVSRAVCHGDLMDIVPLNPRQITVLQICMHACKGRWVSLVAVLKP 588

## RESULT 3

US-10-021-657-4  
; Sequence 4, Application US/10021657  
; Publication No. US20020083483A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMMELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021,657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 158  
; TYPE: PRT  
; ORGANISM: Sorghum sp.  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (1)-(158)  
; OTHER INFORMATION: "Xaa" may be any, other or unknown amino acid  
US-10-021-657-4

Query Match 25.0%; Score 770; DB 13; Length 158;  
Best Local Similarity 94.3%; Pred. No. 3e-72;  
Matches 149; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 87 MPFTSYTYIADPVNVEHLKTNFTNPKGIYVRSYMDVLLGDGIFNADGELWRKORTAS 146  
DB 1 MPFTSYTYIADPVNVEHLKTNFTNPKGIYVRSYMDVLLGDGIFNADGELWRKORTAS 60  
QY 147 FEFASKNLRDFAIVFREYSKLGSLSQASKAGKVVDVMOELYMRTLDSICKVGFVRI 206  
DB 61 FEFASKNLRDFAIVFREYSKLGSLSQASKAGKVVDVMOELYMRTLDSICKVGFVRI 206  
QY 207 GTLSPLDLPENSFAQAFDAANIITLRFIDPLWRIKRF 244  
DB 121 GTLSPLDLPENSFAQAFDAANIITLRFIDPLWRIKRF 158

## RESULT 4

US-10-412-000-4  
; Sequence 4, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:

APPLICANT: ALBERTSEN, MARC C.  
APPLICANT: FOX, TIM  
APPLICANT: HUFFMAN, GARY  
APPLICANT: TRIMMELL, MARY  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
TITLE OF INVENTION: METHOD OF USING SAME  
FILE REFERENCE: 1148  
CURRENT APPLICATION NUMBER: US/10/412,000  
CURRENT FILING DATE: 2003-04-11  
PRIOR APPLICATION NUMBER: US/09/670,153  
PRIOR FILING DATE: 2001-06-11  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 158  
TYPE: PRT  
ORGANISM: Sorghum sp.  
FEATURE:  
NAME/KEY: MOD RES  
LOCATION: (1)-(158)  
OTHER INFORMATION: "Xaa" may be any, other or unknown amino acid  
US-10-412-000-4

Query Match 25.0%; Score 770; DB 14; Length 158;  
Best Local Similarity 94.3%; Pred. No. 3e-72;  
Matches 149; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 87 MPFTSYTIADPVNVEHLKNTNFTNPKGIVYRSDVLLGGIPNADGELWRKORTAS 146  
DB 1 MPFTSYTIADPVNVEHLKNTNFTNPKGIVYRSDVLLGGIPNADGELWRKORTAS 60

QY 147 FEFASKNLRDPSAIFREYSKLSGILSOAKAGKQVDMQELYMWTLDSCIKYGFYEI 206  
DB 61 FEFASKNLRDPSANFREYSKLSGILSOAKAGKQVDMQELYMWTLDSCIKYGFYEI 120

QY 207 GTLSPLDPSNFAQAFDAANIITLRFIDPLWRKRF 244  
DB 121 GTLSPLDPSNFXQAFDAANIITLRFIDPLWRKRF 158

RESULT 5  
US-10-369-493-3749  
Sequence 3749, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 3749  
LENGTH: 492  
TYPE: PRT  
ORGANISM: Neurospora crassa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(492)  
OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-3749

Query Match 13.0%; Score 401; DB 15; Length 492;  
Best Local Similarity 27.2%; Pred. No. 9.7e-33;  
Matches 153; Conservative 71; Mismatches 177; Indels 162; Gaps 26;

QY 66 YHRMDLVGLSRRTVTVDMPFTSYTIADPVNVEHLKNTNFTNPKGIVYRSDVLL 125

DB 1 WQMFRRSIGGTGVRVLVGHRIIFTS-----EPENKAILATQEDYKGGFHOEWKDF 55  
QY 126 LGDGIFNADGELMRKORTASFEFASKNLRDPSA-----IVFREYSL--KLSG---ILS 174  
DB 56 LGDSIFITDGLMHASQLIRPQFIKNRVSDLCQCFENHMQMLFRTIANGGALNGEDQWVD 115  
QY 175 QASKAGKQVDMQELYMWTLDSCIKYGFYEIIGTLDPLD-----ENS-----F 218  
DB 116 MEAGNGKPVDSIDLPFRYTLDAATDFLLGDKDIKSLRFVLPNASFPICXNATENSTPVQPF 175  
QY 219 AQAQF-DAANIITLRFIDPLWRI--KRFEHVSEALLAQSIKLYDE-FTYSVIRRRKAEI 274  
DB 176 ADAFOEVRQVQIVARAGPLNRFVPKTFWEG-----LKVIDETINFIDRLALRDE 227  
QY 275 VEVRASGKQKMKHDIILSRFIELGEAGDGG-----GGFGDDKS-LRDVVNLFVIAGRD 326  
DB 228 BELASKSK-----GDEGYTFELHALAGFTKNRQVLHDQLMAVLLAGRD 269  
QY 327 TTATILSWFTHMAMSHDPVAEKLRRRELCAFEARAREEGVTILVLCGGADADDKAPARVA 386  
DB 270 TTACTLSWAIYELARHPEAVAKRAEI-----LSVVG-----PDRA----- 305  
QY 387 QFAGLLTYDSLGLVLYHACVTETRLRYPAVP--QDP-----KQIL 425  
DB 306 -----PYDDLSKMYLQNVNTEILRLYPVPVFKQVETSFSTRSCFLTFIVISVRVAL 360  
QY 426 EDDVLPDGTQVRAGG-----MVTYVPYSMGEMEYNWG-----PDAASFRPE 466  
DB 361 KDTTLP-----RGGGPDGSGQPIVILKDTVPVGYSLAMQRRPDLVPVSEKFPDVMESPD 415  
QY 467 RWINEDGAFRNASPFKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFELLE----- 520  
DB 416 RWFH-----WQPKWQVPIPFNGGPRICITGQOQFALTENGVTILRFQRYD-RWVSMDI 469  
QY 521 -GHPVQVRMMT--ILSMAGHLKV 540  
DB 470 GQKP---RMKTDIVLMPGDGVKV 489

RESULT 6  
US-09-911-781-32  
Sequence 32, Application US/09911781  
Patent No. US20020034788A1  
GENERAL INFORMATION:  
APPLICANT: Craft, David L.  
APPLICANT: Wilson, C. Ron  
APPLICANT: Eirich, Dudley  
APPLICANT: Zhang, Yeyan  
TITLE OF INVENTION: USE OF CYP52A2A PROMOTER TO INCREASE GENE EXPRESSION IN YEAST  
FILE REFERENCE: U0012 OS/OAAP (1010-49)  
CURRENT APPLICATION NUMBER: US/09/911.781  
CURRENT FILING DATE: 2001-07-24  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 32  
LENGTH: 517  
TYPE: PRT  
ORGANISM: Candida tropicalis  
US-09-911-781-32

Query Match 12.9%; Score 396; DB 9; Length 517;  
Best Local Similarity 26.6%; Pred. No. 3.5e-32;  
Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;

QY 93 TYIA-----DPNVNVEHLKNTNFTNPKGIVYRSDVLLGGIPNADGELWRKQ 141  
DB 84 TYVSLFGTRIVVTYKDPENIKAILATQFGDPSLGRHTLFXP-LLGQIFILDEGKWS 142  
QY 142 RKTASFEFASKNLRDPSA-----VPREYSKLSGILSOAKAGKQVDMQELYMWTLD 196  
DB 143 RMLRPPQFAREQVAHVTSLEPHFOLLKKHLK-----HKGEYFDIQELFFRFTVDS 193



[illegible]

RESULT 10  
 US-10-138-905-100  
 ; Sequence 100, Application US/10138905  
 ; Publication No. US20030068800A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wilson, Ron C.  
 ; APPLICANT: Craft, David L.  
 ; APPLICANT: Eirich, Dudley  
 ; APPLICANT: Eshoo, Mark  
 ; APPLICANT: Madduri, Krishna M.  
 ; APPLICANT: Cornett, Cathy A.  
 ; APPLICANT: Brenner, Alfred A.  
 ; APPLICANT: Tang, Maria  
 ; APPLICANT: Loper, John C.  
 ; APPLICANT: Gleeson, Martin  
 ; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME  
 ; TITLE OF INVENTION: P450 OXIDOREDUCTASE  
 ; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF  
 ; TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
 ; FILE REFERENCE: 1010-16  
 ; CURRENT APPLICATION NUMBER: US/10/138,905  
 ; CURRENT FILING DATE: 2002-05-03  
 ; PRIOR APPLICATION NUMBER: US/09/976,800  
 ; PRIOR FILING DATE: 2001-10-12  
 ; NUMBER OF SEQ ID NOS: 118  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 100  
 ; LENGTH: 517  
 ; TYPE: PRT  
 ; ORGANISM: CANDIDATROPICALIS  
 US-10-138-905-100

[illegible]



Db 194 ATEFLGESVHSLKDESIGINQDDIDFAGRKDFAESFNKAQEYLAIRTL-----VQTFYW 248  
QY 246 VGSEALLAQSILKLVDEFT--YSVIRRRKAEIVFV-RASG-----KQKMKHDLISRFI 295  
Db 249 LVNNKEFRDCTKLHKFTNYVQKALDASPEELEKQSGVVFYELVKQTR-----298  
QY 296 ELGEAGDDGGGDDKSLRDVLFVFIAGRDITATTLSWFTMHAMSHPDVAEKLRELCA 355  
Db 299 -----DENVLRDQSLNILLAGRDITATTLSFAVFLARHPETWAKLREI-- 343  
QY 356 FEARAREEGVTILVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLAHCVTETLRLYP 415  
Db 344 -----EQQFGL-----GED-----SRVEE-----ITFSLKRCYELKAFINETLRIYP 381  
QY 416 AVPDQPKGILEDDVLP-----DGTG-----VRAGMVTYVPYSGMRMEYNNWGPDAASFRPER 467  
Db 382 SVERNPRIATKNTLPRGGSGDGTSPILQKGAVSYGINSHTLDPVYVGPDAAEFRPER 441  
QY 468 WINEDGAFRNAPFKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGHVPQYR 527  
Db 442 WPEPS---TKLGLWYLPFNGGPRICLGQOFA-----470  
QY 528 MMTILSMAHGLKVRVSRAVCHGDLMDIV--PLNPRQITILVLQ 568  
Db 471 ----LTEAGYVLVRLVQEFSHVRLDPDEVYPPKRLTNLTWCLQ 509

## RESULT 13

US-10-139-218-100  
; Sequence 100, Application US/10139218  
; Publication No. US2003015306A1  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Ron C.  
; APPLICANT: Craft, David L.  
; APPLICANT: Eirich, Dudley  
; APPLICANT: Esboo, Mark  
; APPLICANT: Madduri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME  
; TITLE OF INVENTION: P450 OXIDOREDUCTASE  
; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF C  
; TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
; FILE REFERENCE: 1010-16  
; CURRENT APPLICATION NUMBER: US/10/139,218  
; PRIOR FILING DATE: 2002-05-03  
; PRIOR FILING DATE: US/09/976,800  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 100  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: CANDIDATROPICALIS  
US-10-139-218-100

Query Match 12.9%; Score 396; DB 14; Length 517;  
Best Local Similarity 26.6%; Pred. No. 3.5e-32;  
Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;  
QY 93 TYIA-----DPNVVHVLKTNFTNPKGIVYRSYMDVLLGDGFNADGELWRKQ 141  
Db 84 TVVSILFGTRIVVTKDPENIKAILATQFGDFSLGKRHTLFKP-LLGDGIFTLDDGEGWKHS 142  
QY 142 RKTASFEFASKNLRDFAI-----VFREYSILKSLGILSOASKAGKVVDMQELYMRMTLDS 196  
Db 143 RAMLRPQFAREQVAHVTSLEPHQLKKHLK-----HKGEYFDIQELFFRFTVDS 193  
QY 197 ICKVGVGEIGTSLPD-----LPENSFAQAFDAANIITLRFIDPLWRIKRFPH 245

Db 194 ATEFLGESVHSLKDESIGINQDDIDFAGRKDFAESFNKAQEYLAIRTL-----VQTFYW 248  
QY 246 VGSEALLAQSILKLVDEFT--YSVIRRRKAEIVFV-RASG-----KQKMKHDLISRFI 295  
Db 249 LVNNKEFRDCTKLHKFTNYVQKALDASPEELEKQSGVVFYELVKQTR-----298  
QY 296 ELGEAGDDGGGDDKSLRDVLFVFIAGRDITATTLSWFTMHAMSHPDVAEKLRELCA 355  
Db 299 -----DENVLRDQSLNILLAGRDITATTLSFAVFLARHPETWAKLREI-- 343  
QY 356 FEARAREEGVTILVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLAHCVTETLRLYP 415  
Db 344 -----EQQFGL-----GED-----SRVEE-----ITFSLKRCYELKAFINETLRIYP 381  
QY 416 AVPDQPKGILEDDVLP-----DGTG-----VRAGMVTYVPYSGMRMEYNNWGPDAASFRPER 467  
Db 382 SVERNPRIATKNTLPRGGSGDGTSPILQKGAVSYGINSHTLDPVYVGPDAAEFRPER 441  
QY 468 WINEDGAFRNAPFKFTAFQAGPRICLGKDSAYLQMKMALAILFRFYSFRLLEGHVPQYR 527  
Db 442 WPEPS---TKLGLWYLPFNGGPRICLGQOFA-----470  
QY 528 MMTILSMAHGLKVRVSRAVCHGDLMDIV--PLNPRQITILVLQ 568  
Db 471 ----LTEAGYVLVRLVQEFSHVRLDPDEVYPPKRLTNLTWCLQ 509

## RESULT 14

US-10-400-902-32  
; Sequence 32, Application US/10400902  
; Publication No. US20030175896A1  
; GENERAL INFORMATION:  
; APPLICANT: Craft, David L.  
; APPLICANT: Wilson, C. Ron  
; APPLICANT: Eirich, Dudley  
; APPLICANT: Zhang, Yeyan  
; TITLE OF INVENTION: USE OF CYP52A2A PROMOTER TO INCREASE GENE EXPRESSION IN YEAST  
; FILE REFERENCE: U0012 OS/OAAP (1010-49)  
; CURRENT APPLICATION NUMBER: US/10/400,902  
; CURRENT FILING DATE: 2003-03-31  
; PRIOR APPLICATION NUMBER: US/09/911,781  
; PRIOR FILING DATE: 2001-07-24  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 32  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Candida tropicalis  
US-10-400-902-32

Query Match 12.9%; Score 396; DB 14; Length 517;  
Best Local Similarity 26.6%; Pred. No. 3.5e-32;  
Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;  
QY 93 TYIA-----DPNVVHVLKTNFTNPKGIVYRSYMDVLLGDGFNADGELWRKQ 141  
Db 84 TVVSILFGTRIVVTKDPENIKAILATQFGDFSLGKRHTLFKP-LLGDGIFTLDDGEGWKHS 142  
QY 142 RKTASFEFASKNLRDFAI-----VFREYSILKSLGILSOASKAGKVVDMQELYMRMTLDS 196  
Db 143 RAMLRPQFAREQVAHVTSLEPHQLKKHLK-----HKGEYFDIQELFFRFTVDS 193  
QY 197 ICKVGVGEIGTSLPD-----LPENSFAQAFDAANIITLRFIDPLWRIKRFPH 245  
Db 194 ATEFLGESVHSLKDESIGINQDDIDFAGRKDFAESFNKAQEYLAIRTL-----VQTFYW 248  
QY 246 VGSEALLAQSILKLVDEFT--YSVIRRRKAEIVFV-RASG-----KQKMKHDLISRFI 295  
Db 249 LVNNKEFRDCTKLHKFTNYVQKALDASPEELEKQSGVVFYELVKQTR-----298  
QY 296 ELGEAGDDGGGDDKSLRDVLFVFIAGRDITATTLSWFTMHAMSHPDVAEKLRELCA 355  
Db 299 -----DENVLRDQSLNILLAGRDITATTLSFAVFLARHPETWAKLREI-- 343



QY 356 FEAEARAREGVTLLVLCGGADADDKAFARVAQAQFAGLLTYDSLGKLVYLHACVTETRLYP 415  
Db 344 -----EQQFGL-----GED-----SRVEE-----ITFESLKRCEYLKAFINETLRIYP 381  
QY 416 AVPODPKGILEDVLP-----DGTK-----VRAGMVTYVPYSMGRMEYNWGPDAASFRPER 467  
Db 382 SVPRNFRIATKNTTLPGRGGSDGTSPILIQKEAVSYGINSTHLDPPVYIGPDAAEFPER 441  
QY 468 WINEDGAFRNASPFKFTAFQAGPRICLKGDSAYLQMKWALAILFRFYSFRLLLEGHPVQYR 527  
Db 442 WFEPS-----TKKLWAYLPFNGGPRICLGGQFA----- 470  
QY 528 MMTILSMAGHLKVRVSRVAVCHGDLMDIV--PLNPRQITVLVQ 568  
Db 471 ----LTEAGYVLRLVQEFHSHVRLDPDEVYPPKRLTNLTMCILQ 509

Search completed: March 2, 2004, 15:34:53  
Job time : 40 secs

QY 356 FEAEARAREGVTLLVLCGGADADDKAFARVAQAQFAGLLTYDSLGKLVYLHACVTETRLYP 415  
Db 344 -----EQQFGL-----GED-----SRVEE-----ITFESLKRCEYLKAFINETLRIYP 381  
QY 416 AVPODPKGILEDVLP-----DGTK-----VRAGMVTYVPYSMGRMEYNWGPDAASFRPER 467  
Db 382 SVPRNFRIATKNTTLPGRGGSDGTSPILIQKEAVSYGINSTHLDPPVYIGPDAAEFPER 441  
QY 468 WINEDGAFRNASPFKFTAFQAGPRICLKGDSAYLQMKWALAILFRFYSFRLLLEGHPVQYR 527  
Db 442 WFEPS-----TKKLWAYLPFNGGPRICLGGQFA----- 470  
QY 528 MMTILSMAGHLKVRVSRVAVCHGDLMDIV--PLNPRQITVLVQ 568  
Db 471 ----LTEAGYVLRLVQEFHSHVRLDPDEVYPPKRLTNLTMCILQ 509

RESULT 15  
US-10-405-660-100  
; Sequence 100, Application US/10405660  
; Publication No. US20030186411A1  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Ron C.  
; APPLICANT: Craft, David L.  
; APPLICANT: Eirich, Dudley  
; APPLICANT: Eshoo, Mark  
; APPLICANT: Madduri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450 OXIDOREDUCTASE  
; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF CYTOCHROME P450  
; TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
; FILE REFERENCE: 1010-16  
; CURRENT APPLICATION NUMBER: US/10/405,660  
; CURRENT FILING DATE: 2003-04-02  
; PRIOR APPLICATION NUMBER: US/09/976,800  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 100  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: CANDIDATROPICALIS  
US-10-405-660-100

Query Match 12.9%; Score 396; DB 14; Length 517;  
Best Local Similarity 26.6%; Pred. No. 3.5e-32;  
Matches 139; Conservative 78; Mismatches 162; Indels 144; Gaps 19;

QY 93 TYIA-----DPNVHVLKNTFTNPKGIVYRSYMDVLLGDGIFNADGELWRKQ 141  
Db 84 TYVILFGRIVTVTKDENIKAILATQFGPSLGKRLFKP-LLGDGIFTLDEGWKHS 142  
QY 142 RKTASFEFASKNLRDESAI-----VFREYSLKLSGILSQASKAGKVVDMQELYNRNLDS 196  
Db 143 RAMLRPOQFAREQVAHVITSLBPHFQLKKHLK-----HKGEYFDIQELFFRFTVDS 193  
QY 197 ICKVGFGEIGTLSPD-----LPNSFAQAADAANIITLRFIDPLWRIKKEFEH 245  
Db 194 ATEFLFGESVHSLKDESGINQDDIDFAGRXDFAESFNKAQEYLAIRTL-----VQTFYW 248  
QY 246 VGSEALLAQSTKLYDEFT--YSVIRRRKAEIIEV-RASG-----KQEKMKHDLISRFI 295  
Db 249 LVNNKGFDCVKLVHKFTNYTVQKALDASPEELEKQSGYVFLVYLVQTR----- 298  
QY 296 ELGEAGDGGGFGDDKSLRVIAGRDRTTATTLTSWFTMHAMSHPDVAEKLRLLECA 355  
Db 299 -----DENVLRDQSLNILLAGRDTTAGLLSFVAFELARHPEIWAUKLREEI-- 343

Sequence 28, Appl  
Sequence 28, Appl  
Sequence 30, Appl  
Sequence 20, Appl  
Sequence 4, Appl  
Sequence 12, Appl  
Sequence 68, Appl  
Sequence 2, Appl  
Sequence 6, Appl  
Sequence 405, Appl  
Sequence 4, Appl  
Sequence 70, Appl  
Sequence 4, Appl  
Sequence 6, Appl  
Sequence 4, Appl  
Sequence 9, Appl  
Sequence 8, Appl  
Sequence 4, Appl

## ATTACHMENTS

```

RESULT 1
; US-09-158-767-19
; Sequence 19, Application US/09158767A
; Patent No. 6180363
; GENERAL INFORMATION:
; APPLICANT: Batard, Yannick
; APPLICANT: Durst, Francis
; APPLICANT: Schalk, Michel
; APPLICANT: Werck-Reichhart, Daniele
; TITLE OF INVENTION: RECODING OF DNA SEQUENCES PERMITTING
; TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST
; FILE REFERENCE: A32000
; CURRENT APPLICATION NUMBER: US/09/158,767A
; CURRENT FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: FR 97-12094
; EARLIER FILING DATE: 1997-09-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Fast-SEQ for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Altered sequences
; US-09-158-767-19

```

Query Match	31.7%;	Score 976;	DB 3;	Length 541;
Best Local Similarity	39.9%;	Pred. No. 5.5e-100;		
Matches	222;	Conservative 82;	Mismatches 190;	Indels 62; Gaps 13;
QY	27	TALLVLSWILVQRWSL-RKQKGRSPWPVIGATVEQLNRYHRMHDWLGVYLSR-----HR	81	
DB	10	VSAVAAYMAWF-----WRMSRGLGRPRVPLVGLSLPGLVQHAEDMHEWIAAGNLRRAAGTYQ	65	
QY	82	TVTVDMPTFS-----YTIYADPNVYEHVLKTNFTNPKGIVYRSYMDVLGGINFADG	136	
DB	66	TCICFVPGVAREGGIVTVCTDPRNLEHVLUKAFADNYPKGPFVHGVFPRDLGGINFNSDGD	125	
QY	137	LWRQKRTASFFASKNLRD-PSAIVFREYSJLKGILSOAKGKGVMDQMELYMRNTLD	195	
DB	126	TWLAQRKTAALFETRTLTATMSRWVYSRSHGRLLPLPILADAAGKAKQVLODILLRLTFD	185	
QY	196	STCKVGFGEVIGTLPDLPENSAFADAANIITLRFIDP--LWRIKPFHVGVSSALLA	253	
DB	186	NTCGLAFGKDPVLAQGLPENEFASAFDRATATLNRFIPFPFELWRCCKWLGLGMETTLT	245	
QY	254	QSTKLVDFTYSVIRRRKAEIVFVSASGKQEKMKHDILSRFIELGAGDDGGGFGDDKSL	313	
DB	245	SNWVHVQVIAVYIKRKLLEAAGNGKCDTAAPHDOLLSRFMKKGYSDB-----SL	297	

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

Run on: March 2, 2004, 15:26:34 ; Search time 22 Seconds  
(without alignments)  
1379.921 Million cell updates/sec

Title: US-10-021-657-2  
 Perfect score: 3079  
 Sequence: 1 EFGTREHILTPATSPPEPL.....ICMHA CKGRWVSLVAVLKP 589

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:	389414
--	--------

Minimum DB seq length: 0

[illegible]

Post-processing: Minimum Match 0%

Maximum Match 100%  
 Missing files 45 summarized

```
Database : Issued Patents Aa:*
1: /cgn2_6/pdata2/iaa/5A COMB.pep:*
2: /cgn2_6/pdata2/iaa/5B COMB.pep:*
3: /cgn2_6/pdata2/iaa/6A COMB.pep:*
4: /cgn2_6/pdata2/iaa/6B COMB.pep:*
5: /cgn2_6/pdata2/iaa/PCTUS COMB.pep:*
6: /cgn2_6/pdata2/iaa/backfiles1.pep:*
```

ppred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	976	31.7	541	3	US-09-158-767-19
2	976	31.7	541	3	US-09-158-767-20
3	396	12.9	517	4	US-09-302-620B-100
4	396	12.9	517	4	US-09-302-620B-101
5	395	12.8	517	4	US-09-302-620B-97
6	393	12.8	522	4	US-09-302-620B-95
7	389	12.6	523	4	US-09-302-620B-96
8	387	12.6	522	4	US-09-302-620B-98
9	383.5	12.5	540	4	US-09-302-620B-98
10	381	12.4	512	4	US-09-302-620B-102
11	380	12.3	512	4	US-09-302-620B-103
12	379	12.3	540	4	US-09-302-620B-99
13	367	11.9	576	3	US-08-948-564-16
14	353	11.5	499	4	US-09-302-620B-104
15	353	11.5	524	4	US-09-976-594-533
16	342	11.1	504	5	US-08-457-274A-25
17	342	11.1	504	5	PC7-US95-05758-25
18	334.5	10.9	557	4	US-09-518-386B-1
19	334.5	10.9	557	4	US-09-518-386B-3
20	329.5	10.7	507	1	US-08-457-274A-23
21	329.5	10.7	507	5	PC7-US95-05758-23
22	328	10.7	510	4	US-09-852-067-2
23	319	10.4	503	4	US-09-144-367-2
24	313.5	10.2	489	4	US-09-852-067-4
25	311	10.1	520	4	US-09-527-073-2
26	306	9.9	517	5	US-08-457-274A-2
27	305	9.9	517	5	PC7-US95-05758-2



QY 356 FEARAREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGLVYLHACVTTETLRYP 415  
Db 344 -----EQOFGI-----GED-----SRVEE-----ITFESLKRCEYLKAFNLNTRIYP 381  
QY 416 AVPDQPKGILEDDVLP-----DGTK-----VRAGGWTVVPYSGMRMEYNWGPDAASFRPER 467  
Db 382 SVPRNFRIATKNTTLFRGGSDGTSPILQKGEAVSYGINSTHLDPVYVYGPDAASFRPER 441  
QY 468 WINEDGAFRNAPSKFTAFQAGPRICLGDSDAYLQMKVALAILFRFYSFRLDEGHPVQYR 527  
Db 442 WFEPSS-----TKLWYVLPFNGGPRICLQGOFA----- 470  
QY 528 MMTILSWAHLKVRVSRAVCHGLDMDIV--PLNPROITVLQ 568  
Db 471 ----LTEAGYVLVRLVQEFSHVRLDPEVYPPKRLTNLTWCLQ 509

## RESULT 4

US-09-911-781-32  
; Sequence 32, Application US/09911781  
; Patent No. 6673613  
; GENERAL INFORMATION:  
; APPLICANT: Craft, David L.  
; APPLICANT: Wilson, C. Ron  
; APPLICANT: Erich, Dudley  
; APPLICANT: Zhang, Yeyan  
; TITLE OF INVENTION: USE OF CYP52A2A PROMOTER TO INCREASE GENE EXPRESSION IN YEAST  
; FILE REFERENCE: U0012 OS/CAAP (1010-49)  
; CURRENT APPLICATION NUMBER: US/09/911,781  
; CURRENT FILING DATE: 2001-07-24  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 32  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Candida tropicalis  
US-09-911-781-32

Query Match 12.9%; Score 396; DB 4; Length 517;  
Best Local Similarity 26.6%; Pred. No. 4.7e-35;  
Matches 139; Conservative 78; Mismatches 163; Indels 144; Gaps 19;

QY 93 TYIA-----DPNVVHVLKNTFTNPKGIVYRSYMDVLLGDIQFADGELARKQ 141  
Db 84 TVSILFETRIWTKDPENIKAILATQGFSLGKRHLFKP-LIGDGIFTLDGEGWKS 142  
QY 142 RKTASFEFASKNLRFSAL-----VREYSKLGLSGLSKAGKAGVDMQELVYRMTLDS 196  
Db 143 RMLRPQFARQVAVHTSLEPHFQLKXHLK-----HKGYFDIQELFFRFTVDS 193  
QY 197 ICKVGFVEIGTLSPD-----LPENSFAQAFDAANIITLRFIDPLWRIKRFPHVSEALLA 245  
Db 194 ATEFLGSGSVHSLKDESIGNQDDIDFAGRKDFAESFNKAQBYLAIRTL-----VQTFY 248  
QY 246 VGSEALLAQSKLVDEFT--YSVIRRKAEIVR-ASG-----KQKMKHDLISRFI 295  
Db 249 LVNNKFEKDCDKLVHKFTNYVQKALDASPEELKQSGVYFVYELVKQTR----- 298  
QY 296 ELGEAGDDGGGDKSLRDVNLNFVIAGRDITATTLTSWFTWMSHDPDVAEKLREELCA 355  
Db 299 -----DPNVLRDQSLNILLAGRDITAGLLSFVAVFELARHPHIMAKLREEI-- 343  
QY 356 FEARAREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGLVYLHACVTTETLRYP 415  
Db 344 -----EQOFGI-----GED-----SRVEE-----ITFESLKRCEYLKAFNLNTRIYP 381  
QY 416 AVPDQPKGILEDDVLP-----DGTK-----VRAGGWTVVPYSGMRMEYNWGPDAASFRPER 467  
Db 382 SVPRNFRIATKNTTLFRGGSDGTSPILQKGEAVSYGINSTHLDPVYVYGPDAASFRPER 441  
QY 468 WINEDGAFRNAPSKFTAFQAGPRICLGDSDAYLQMKVALAILFRFYSFRLDEGHPVQYR 527  
Db 442 WFEPSS-----TKLWYVLPFNGGPRICLQGOFA----- 470

QY 528 MMTILSWAHLKVRVSRAVCHGLDMDIV--PLNPROITVLQ 568  
Db 471 ----LTEAGYVLVRLVQEFSHVRLDPEVYPPKRLTNLTWCLQ 509

## RESULT 5

US-09-302-620B-101  
; Sequence 101, Application US/09302620B  
; Patent No. 6391420  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, C. Ron  
; APPLICANT: Craft, David L.  
; APPLICANT: Erich, Dudley  
; APPLICANT: Eshoo, Mark  
; APPLICANT: Madduri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
; TITLE OF INVENTION: RELATING THERETO  
; FILE REFERENCE: 1010-16.seq  
; CURRENT APPLICATION NUMBER: US/09/302,620B  
; CURRENT FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 109  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 101  
; LENGTH: 517  
; TYPE: PRT  
; ORGANISM: Candida tropicalis  
US-09-302-620B-101

Query Match 12.8%; Score 395; DB 4; Length 517;  
Best Local Similarity 27.5%; Pred. No. 6.1e-35;  
Matches 132; Conservative 77; Mismatches 171; Indels 100; Gaps 16;

QY 90 TSVTYIADPNVHVLKNTFTNPKGIVYRSYMDVLLGDIQFADGELARKQKTASFEF 149  
Db 92 TKIVVTKDPENIKAILATQGFSLGKRHLFKP-LIGDGIFTLDGEGWKSRLRPFQ 150  
QY 150 ASKNLRDPSAL-----VREYSKLGLSGLSKAGKAGVDMQELVYRMTLDSICKVSGEV 204  
Db 151 AREQVAHVTSLEPHFQLKXHLK-----HKGYFDIQELFFRFTVDSATEFULGE 201  
QY 205 EIGTLSPD-----LPENSFAQAFDAANIITLRFIDPLWRIKRFPHVSEALLA 253  
Db 202 SVHSLKDETICINODDIDFAGRKDFAESFNKAQBYLSIRIL-----VQTFYMLNNKEFR 256  
QY 254 QSIKLVDEFT--YSVIRRKAEIVR-ASG-----KQKMKHDLISRFIELGEAGDD 303  
Db 257 DCTKLHVKFTNYVQKALDASPEELKQSGVYFVYELVKQTR----- 298  
QY 304 GGGGDDKSLRDVNLNFVIAGRDITATTLTSWFTWMSHDPDVAEKLREELCAFEARARE 363  
Db 299 -----DPNVLRDQSLNILLAGRDITAGLLSFVAVFELARHPHIMAKLREEI-- 344  
QY 364 EGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGLVYLHACVTTETLRYPVQDPKG 423  
Db 345 QOFGI-----GED-----SRVEE-----ITFESLKRCEYLKAFNLNTRIYPVPRNFI 389  
QY 424 ILEDDVLP-----DGTK-----VRAGGWTVVPYSGMRMEYNWGPDAASFRPRWINEGAF 475  
Db 390 ATKNTTLFRGGSDGTSPILQKGEVSYGINSTHLDPVYVYGPDAASFRPRWINEGAF 446  
QY 476 RNAPSKFTAFQAGPRICLGDSDAYLQMKVALAILFRFYSFRLDEGHPVQYRMTILSM 534  
Db 447 TRKLWYVLPFNGGPRICLQGOFA----- 506

```
RESULT 6
US-09-302-620B-97
; Sequence 97, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Eirich, Dudley
; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: 1010-16.seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 97
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-97

Query Match 12.8%; Score 393; DB 4; Length 522;
Best Local Similarity 28.4%; Pred. No. 1e-34;
Matches 140; Conservative 73; Mismatches 184; Indels 96; Gaps 20;

QY 77 LSHRTVTVDMPSTYIIA--DPNVHVLKNTFTYKGIYVRSYMDVLGDIAGNAGELWKKORKTASFEFASKNLRD 134
DB LNPDPDPTFTFFESINLSTLEPENIKAILATQFNDPSLG-TRHSHFAPLLGDIQIFLTD 140
QY 135 GELWKKORKTASFEFASKNLRDPSAIVFREYSYKLSGILSQASKA-GKVVDMDQLYRYMT 193
DB 141 GAGWKHSRMLRQFAREQI---SHVKLLPEHMQV--FFKHVRKAQCKTFDIQELFFRLT 195
QY 194 LDISCKVGGVEIGTSLPDLFENSPFAQFADAANIITLRFIDPLWIKRFFHVGSEALLA 253
DB 196 VDSATBFLFGESVESLR---DESIGMSINALDFDGKAGFADAFNYSQNY--LASAVMQ 249
QY 254 QSIKLVDFTYSVIRRRKAEIVFRASGKQKMKHDLISRFIELGE---AGDDGGGF--- 307
DB 250 Q-----IYWLNGKKFK-----ECNAKVHKADYVVSXKALDLTPEQLEKQDGYVFLYE 297
QY 308 -----GDKSLRDVLFVIAGDRTTATTLTWSFTHMAMSHPDVAEKLRLRELCAFEAEAR 362
DB 298 LVKQTRDRQVLRQQLNIMVAGDRTTAGLLSFVFFELARNPEVTNKLREBI----- 348
QY 363 BEGTVLILCGADADDK---AFARVAQAFAGLTYSLSGLKLVILHACVTTETLRLYPVPQ 419
DB 349 -----EDKFGLENARVED-----ISFESLKCEYLKALINETLRLYPSVPQ 390
QY 420 DPKGILEDDVLDPG-----TKVRAGGMVTVYPYSGMGEYVNWGPDAAASFRPERWINE 471
DB 391 NFRVATKNTTLPGGGKDGSLPVLVRGQTVMYGVYAAHNPVAVYKDALEFRPERWFE 450
QY 472 DGAFNRPSPKFTAFQAGPRICLKDSAYLQMKKALAILFRFYSGFRILLE--GH-----PV 524
DB 451 E---TKKLGWAFPLFNGGPRICLQGFALTEAS-----YVTVRILQBFHLSMDPNT 499
QY 525 QY--RMTLSMA 535
DB 500 EYPPRKMSHLTMS 512

RESULT 7
US-09-302-620B-95
```

```
; Sequence 95, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Eirich, Dudley
; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; APPLICANT: Gleeson, Martin
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: 1010-16.seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 95
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-95

Query Match 12.6%; Score 389; DB 4; Length 523;
Best Local Similarity 28.1%; Pred. No. 2.9e-34;
Matches 127; Conservative 60; Mismatches 167; Indels 98; Gaps 16;

QY 97 DPVNVHVLKNTFTYKGIYVRSYMDVLGDIAGNAGELWKKORKTASFEFASKNLRD 156
DB 105 DENIKAVLATQFDFSLG-TRHSHFAPLLGDIQIFLTDGEGWKHGRAMLRPQFADQIGH 163
QY 157 PSA-----IVFREYSYKLSGILSQASKAGKVVDMDQLYRYMTLDSICKVGGVEIGTL--- 209
DB 164 VKALPEPHIQIMAKQIKLN-----QKTFDIQELFFRTVDTATEFLFGESVHSLYDE 215
QY 210 -----SP-DLP-ENSPFAQFADAANIITLRFIDP-----LWRTKRFPHVGESEALLAQSIKL 258
DB 216 KLGIPTPEIPGRENFAAFNVSOHYLATRSYSQTFYFLTNPKFPRDCN-----AKVHHL 270
QY 259 VDEFTYSVIRRRKAEIVFRASG-----KQKMKHDLISRFIELGAGDDGGFGDD 310
DB 271 AKYFVNKALNFTPEBLEEKSGYVFLYELVKQTR-----DP 307
QY 311 KSLRDVLFVIAGDRTTATTLTWSFTHMAMSHPDVAEKLRLRELCAFEAEARAEEGVTLVL 370
DB 308 KVLQDQLNIMVAGDRTTAGLLSFALFELARHPKMWKSLREBI-----EYVFGV--- 356
QY 371 CGGADADDKAFARVAQAFAGLTYSLSGLKLVILHACVTTETLRLYPVPQPKGILEDDVL 430
DB 357 --GED-----SRVEE-----ITFEALKRCEYLKALINETLRLMYSPVPVNFRTATRDITL 403
QY 431 P-----DGKTVRAGGMVTVYPYSGMGEYVNWGPDAAASFRPERWINEDEGAFNRPSPK 482
DB 404 PRGGGANGTDPIYIPKSGTVAVVVYVYKTHRELYEYKGDANDFRPERWFEP-----TKKLGWA 460
QY 483 FTAFQAGPRICLKDSAYLQMKKALAILFRFY 514
DB 461 YVFPNGPRVCLGQGFALTEASVITRLAQM 492

RESULT 8
US-09-302-620B-96
; Sequence 96, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Eirich, Dudley
```

Mon Mar 8 10:13:27 2004

APPLICANT: Eshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.  
APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
TITLE OF INVENTION: RELATING THERETO  
FILE REFERENCE: 1010-16.seq  
CURRENT APPLICATION NUMBER: US/09/302,620B  
CURRENT FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 109  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 96  
LENGTH: 522  
TYPE: PRT  
ORGANISM: Candida tropicalis  
US-09-302-620B-96

Query Match 12.6%; Score 387; DB 4; Length 522;  
Best Local Similarity 27.3%; Pred. No. 4.9e-34;  
Matches 131; Conservative 69; Mismatches 185; Indels 94; Gaps 16;  
QY 68 RMHDLVGLSRHRTVTVDMPFTSYTI--ADPVNVEHVLTNTNYPKGVVRSYMDVL 125  
DB 78 RIHD-----LDRPDITFTFPVFSINLVNTLEPENIKAILATQNDPSLG-TRHSHFAPL 131  
QY 126 LGDGFNADGELWRKQRTASFPASKNLRDPSAI-----VFREYSIKLSGILSQASKAG 180  
DB 132 LGDGFITLDGAGWKHSRMLRPOFARQISHVKLEPHVQVFFKHVKAQ-----G 182  
QY 181 KVDVMQELVMTLDSICKVGFVEIGTSLPDLSPENGFAQAFDAANIITLRFIDPLMRI 240  
DB 183 KTFDIQELFRLTVDSTAFELFGESVSLR-----DESIGSINALDPDGKAGFADAFNS 238  
QY 241 KRFFHVGEALLAQSILKIVDEFTYSVIRRRKAEIVVRAKQKQKMDILSRFELGE- 299  
DB 239 QNY--LASRAVWQ-----LYWLVNGKFKF---ECNAKVKFADYVYNKALDLTPE 284  
QY 300 --AGDGGGF-----GDKSLRDVVLNVFVIAGRTTATLTSWFTMHMNSHPDVAEL 349  
DB 285 QLEKQDGVVFLYELVKQTRDKQVLRDQLLNTMVAAGRTTAGLLSFVFFELARNPEVTNKL 344  
QY 350 RRELCAFEARAEEGVTIVLCGGADADDKAFARVAQFAGLLTVDSLGLKLVYLHACVTE 409  
DB 345 REEI-----EDKFLGENASVED-IGFESLSKCEYLKAVLINE 380  
QY 410 TRLYPVAVPQPKGILEDDVLPDG-----TKVRAGGMVTVYVPSNGRMEYNWGPDA 461  
DB 381 TRLYPSPQNFVRVATKNTTLPRGGGKDLSPVLVRKQTVIYGVYAAHRNPVAVYKDAL 440  
QY 462 SFRERWINEGAFRNASPFKFTAFQAPRICLKDSAYLQMKWALALFLRFYSFRLLE 520  
DB 441 BFRPERWFEPE---TKLWGAFLPENGPPRICLGQQQFALTEAS-----YTVRLLQ 488

RESULT 9  
US-09-302-620B-98  
Sequence 98, Application US/09302620B  
Patent No. 6331420  
GENERAL INFORMATION:  
APPLICANT: Wilson, C. Ron  
APPLICANT: Craft, David L.  
APPLICANT: Eirich, Dudley  
APPLICANT: Eshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.

APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
TITLE OF INVENTION: RELATING THERETO  
FILE REFERENCE: 1010-16.seq  
CURRENT APPLICATION NUMBER: US/09/302,620B  
CURRENT FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 109  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 98  
LENGTH: 540  
TYPE: PRT  
ORGANISM: Candida tropicalis  
US-09-302-620B-98

Query Match 12.5%; Score 383.5; DB 4; Length 540;  
Best Local Similarity 25.5%; Pred. No. 1.3e-33;  
Matches 148; Conservative 87; Mismatches 209; Indels 137; Gaps 23;  
QY 5 REAHLTPA---TPSPFFPLAGBHKYIALLLVLSWILVQRWSLRKQK---PRSWPVIGA 58  
DB 53 RRFHAKPLGNFVRDPTFGIATP-----LLLIYL-----KSGIVMKFAWGL--- 93  
QY 59 TVBQLRNYHRMHDWLVGYLSR---HRTVTVDMPFTSYTIADPVNVEHVLTNTNYPK 115  
DB 94 -----WNNKYIVRDKYKXTGLRIVGLPLIETMDPENIKAVLATQNDPSLG 140  
QY 116 IVRSYMDVLLDGIIFNADGELWRKQRTASFPASKNLRDPSAI-----VFREYSIKLS 170  
DB 141 -TRHDFLYSLDGIITLDGAGWKHSRMLRPOFARQISHVKLEPHVQVFFKHVKAQ--- 197  
QY 171 GILSQASKAGKVDVMQELVMTLDSICKVGF-----VEIGTSLPDLPE---NSF 218  
DB 198 -----HRGQTFDIQELFRLTVDSTAFELFGESABSLRDESIG-LTPTKDFDGRDF 249  
QY 219 AQAFDAANIITLRFIDPLMRIKRFFHVGEALLAQSILKIVDEFTYSVIRRRKAEIVEVR 278  
DB 250 ADAFNSQTVQAYRFL-----LQMYWILNGSFRKSIIVVHKFADHYV-QKALELTD-- 301  
QY 279 ASGKQEKMKHDIILSRFIELGEAGDGGGFDKSLRDVVLNVFVIAGRTTATLTSWFTMH 338  
DB 302 ---DDLQKQDGVVFLYELAKQ-----TRDPKVLRLDQLLNIIVAGRTTAGLLSFVFE 351  
QY 339 AMGHPDVAELKRELCAFEARAEEGVTIVLCGGADADDKAF---ARVAQFAGLLTY 394  
DB 352 LSRNPEVFAKLREBVE-----ENRFGLGEARVEE-----ISF 383  
QY 395 DSLGKLVYLHACVTEITLRLYPVAVPQD-----PKGILEDDVLPDGTKVRAGGMVTV 444  
DB 384 ESLKSCYELKAVINETLRLYPSPHNFVRVATNTTLPRGGGEDGYSP--IWKKGQVVMY 441  
QY 445 VPYSGRMEYNWGPDAASFRPERWINEGAFRNASPFKFTAFQAPRICLKDSAYLQMK 504  
DB 442 TVIATHRDPISYIGADADVFRPERWFEPE---TRKLGMAYVFPNGGPPRICLGQQQFALTEAS 498  
QY 505 MALAILFRFYSFRLLE---GHPVQVRMTTILSMAHGLKVVV 542  
DB 499 YTVRLLIQEFAHLSMDPDTETYPKQLQNTLTLSLFDGADVRM 539

RESULT 10  
US-09-302-620B-102  
Sequence 102, Application US/09302620B  
Patent No. 6331420  
GENERAL INFORMATION:  
APPLICANT: Wilson, C. Ron  
APPLICANT: Craft, David L.  
APPLICANT: Eirich, Dudley  
APPLICANT: Eshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.



; APPLICANT: Tang, Maria  
 ; APPLICANT: Loper, John C.  
 ; APPLICANT: Gleeson, Martin  
 ; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
 ; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
 ; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
 ; TITLE OF INVENTION: RELATING THERETO  
 ; FILE REFERENCE: 1010-16.seq  
 ; CURRENT APPLICATION NUMBER: US/09/302,620B  
 ; CURRENT FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 109  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 102  
 ; LENGTH: 512  
 ; TYPE: PRT  
 ; ORGANISM: Candida tropicalis  
 US-09-302-620B-102

Query Match 12.4%; Score 381; DB 4; Length 512;  
 Best Local Similarity 26.1%; Pred. No. 2.2e-33;  
 Matches 145; Conservative 89; Mismatches 181; Indels 140; Gaps 24;  
 QY 26 YIAL-LLVLSWIL--VORSLRKQKGRSPVPI-----GATVEQLNHYRMDW 72  
 DB 10 YIVLELAIQIVAHVTRNLMKLGAKPFTHVORDGWLGFGRFLKAKSAGRLVDL 69  
 QY 73 LVGLSRHRTVDMPTFSY-----TYIADPNVNEHVILKTNFTNPKGIVYRSMYVLL 126  
 DB 70 IISRFHEDNT-----FSSYAFGNHVFTROPENIKALLATQFGDFSLGSRVK-FFKELL 123  
 QY 127 GDGIFNADGELWRKQRTASFFASKNLRDPSAI-----VFREYSLKSLGILSQASKAGK 181  
 DB 124 GYGIFTLDAEGKWSRAMLRPOFAREQVAHVTSLEPHFOLLKHLK-----HKGE 174  
 QY 182 VDMQELYNMTLDSICKVGF-----VEIGTLPDLPE--NSFAQAFDAANIITL 231  
 DB 175 YFDIQELFFRTVDSATEFLFGESVHSLRDEEIGYDTKDMAEERRKPADAFNKSQVYVAT 234  
 QY 232 RFIDPLMRIRKFFHVGSEALLAQSIKLVDEFT--YSVIRRRKAEIIVEVRASG-----281  
 DB 235 RV-----ALONTLYWLVNNKEFCENDIVHKFTNYYVQKALDTPPELEKQGGVFIYELV 289  
 QY 282 KQEKMKHILSRFIELGEAGDGGGDKSLRDVVLNFIAGROTTATTLTSWFTHMAMS 341  
 DB 290 KQTR-----DPKVLDDQSLNILLAGROTTAGLLSFAVFEAR 326  
 QY 342 HPDVAEKRLRELCAFEAREBEGVTLVLCGGADADDKAFARVAOFAGLITVDSLGKLV 401  
 DB 327 NPHIWAQLREEI-----EQOFGI-----GED-----SRVEE-----ITFESLKRCE 362  
 QY 402 YLHACVTTETRLYPVAPQDPKGILEDDVLDPGTVKVRAG-----MVTYVPYSMGR 451  
 DB 363 YLKAVLNELTLRHESVPRNARPAIKDTTLP-----RGGPGNGKDPILIRKDEVVQYSISA 417  
 QY 452 MEYN-----WGPDAASFRPERWINEGDAFRNASPFKFTAFQAGPRICLGKDSAYLQMKWALA 508  
 DB 418 TQTNPAFYGADAADFRPERWF--EPSTRNLG-WAFLPFGNGGPRICLGQOQFALTEAGYVLV 474  
 QY 509 ILFRFYFRLLGHP 523  
 DB 475 -----RLVQEPF 481

RESULT 11  
 US-09-302-620B-103  
 ; Sequence 103, Application US/09302620B  
 ; Patent No. 6331420  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Willson, C. Ron  
 ; APPLICANT: Craft, David L.  
 ; APPLICANT: Birch, Dudley  
 ; APPLICANT: Eshoo, Mark  
 ; APPLICANT: Madduri, Krishna M.

; APPLICANT: Cornett, Cathy A.  
 ; APPLICANT: Brenner, Alfred A.  
 ; APPLICANT: Tang, Maria  
 ; APPLICANT: Loper, John C.  
 ; APPLICANT: Gleeson, Martin  
 ; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
 ; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
 ; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
 ; TITLE OF INVENTION: RELATING THERETO  
 ; FILE REFERENCE: 1010-16.seq  
 ; CURRENT APPLICATION NUMBER: US/09/302,620B  
 ; CURRENT FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 109  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 103  
 ; LENGTH: 512  
 ; TYPE: PRT  
 ; ORGANISM: Candida tropicalis  
 US-09-302-620B-103

Query Match 12.3%; Score 380; DB 4; Length 512;  
 Best Local Similarity 26.1%; Pred. No. 2.9e-33;  
 Matches 145; Conservative 87; Mismatches 183; Indels 140; Gaps 24;  
 QY 26 YIAL-LLVLSWILVORWS--LRKQKGRSPVPI-----GATVEQLNHYRMDW 72  
 DB 10 YIVLPLLVITKQIVAHARTNLMKLGAKPFTHVOLDGWFGRFLKAKSAGRQVDL 69  
 QY 73 LVGLSRHRTVDMPTFSY-----TYIADPNVNEHVILKTNFTNPKGIVYRSMYVLL 126  
 DB 70 IISRFHEDNT-----FSSYAFGNHVFTROPENIKALLATQFGDFSLGSRVK-FFKPLL 123  
 QY 127 GDGIFNADGELWRKQRTASFFASKNLRDPSAI-----VFREYSLKSLGILSQASKAGK 181  
 DB 124 GYGIFTLDDGCKWSRAMLRPOFAREQVAHVTSLEPHFOLLKHLK-----HKGE 174  
 QY 182 VDMQELYNMTLDSICKVGF-----VEIGTLPDLPE--NSFAQAFDAANIITL 231  
 DB 175 YFDIQELFFRTVDSATEFLFGESVHSLRDEEIGYDTKDMAEERRKPADAFNKSQVILST 234  
 QY 232 RFIDPLMRIRKFFHVGSEALLAQSIKLVDEFT--YSVIRRRKAEIIVEVRASG-----281  
 DB 235 RV-----ALQTLVWLVNNKEFCENDIVHKFTNYYVQKALDTPPELEKQGGVFIYELA 289  
 QY 282 KQEKMKHILSRFIELGEAGDGGGDKSLRDVVLNFIAGROTTATTLTSWFTHMAMS 341  
 DB 290 KQTR-----DPNVLDDQSLNILLAGROTTAGLLSFAVFEAR 326  
 QY 342 HPDVAEKRLRELCAFEAREBEGVTLVLCGGADADDKAFARVAOFAGLITVDSLGKLV 401  
 DB 327 NPHIWAQLREEI-----ESHEGL-----GED-----SRVEE-----ITFESLKRCE 362  
 QY 402 YLHACVTTETRLYPVAPQDPKGILEDDVLDPGTVKVRAG-----MVTYVPYSMGR 451  
 DB 363 YLKAVLNELTLRHESVPRNARPAIKDTTLP-----RGGPGNGKDPILIRKDEVVQYSISA 417  
 QY 452 MEYN-----WGPDAASFRPERWINEGDAFRNASPFKFTAFQAGPRICLGKDSAYLQMKWALA 508  
 DB 418 TQTNPAFYGADAADFRPERWF--EPSTRNLG-WAFLPFGNGGPRICLGQOQFALTEAGYVLV 474  
 QY 509 ILFRFYFRLLGHP 523  
 DB 475 -----RLVQEPF 481

RESULT 12  
 US-09-302-620B-99  
 ; Sequence 99, Application US/09302620B  
 ; Patent No. 6331420  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Willson, C. Ron  
 ; APPLICANT: Craft, David L.  
 ; APPLICANT: Birch, Dudley

APPLICANT: Bshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.  
APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
TITLE OF INVENTION: RELATING THERETO  
FILE REFERENCE: 1010-16:seq  
CURRENT APPLICATION NUMBER: US/09/302,620B  
CURRENT FILING DATE: 1999-04-30  
NUMBER OF SEQ ID NOS: 109  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 99  
LENGTH: 540  
TYPE: PRT  
ORGANISM: Candida tropicalis  
US-09-302-620B-99

Query Match 12.3%; Score 379; DB 4; Length 540;  
Best Local Similarity 26.4%; Pred. No. 4,1e-33;  
Matches 126; Conservative 74; Mismatches 187; Indels 90; Gaps 14;

QY 97 DPVNVHVLKTNFTNPKGIYRSYMDVLLGDGIFNADGELMRKQRTASFEFASKNLRD 156  
Db 122 DPENIKAVLATQNDRLSG-TRHDFLYSLDGDGIFNADGELMRKQRTASFEFASKNLRD 180  
QY 157 FSAI-----VFREYSLKLSGLSQAAGKVVDMQELMYMTLDSICKVGFVEIGTSLP 211  
Db 181 VKLLEPHVQVFFKHVRK-----HRGQTDFDIQELFRLLTVDSATEFFLFGSABSLRD 231  
QY 212 DL-----PENSEFAAFDAANIIITLRFIDPLWRIKRPVHVGSEALLAQSIKLV 260  
Db 232 DSVGLTPTTKDFEGRGDFADAFNSQTYQARFL-----LQOMWILNGAEFRKSIALVH 286  
QY 261 EFTYSVIRRAKAEVVEVRASGQKMKHDLISRFIEIGAGDDGGGFGDDKSLRDVVLNF 320  
Db 287 KFADHYV-QKALETLD-----DLQKQGVVFLYELAKQ-----TRDPKVLRLDQLLNI 333  
QY 321 VIAGRTTATTLTWNFMHMSHPDVAEKLRELCFAFAERAREBGVTILVLCGGADADKA 380  
Db 334 LVAGRTTATTLTWNFMHMSHPDVAEKLRELCFAFAERAREBGVTILVLCGGADADKA 380  
QY 381 FA-----ARVAQFAGLLTVDSLGLVYLHACVTEITRLYPAPQDPKGILEDVLPDG--- 433  
Db 370 FGLGEEARVES-----ISFESLKSCEYLKAVINEALRLYPSVPHNFRVATRNITLPRGGK 425  
QY 434 -----TKVRAGGMVTVYPYSMGRMEYNWGPDAAGFRPERWINEGAFRNAGPFKTAPOA 488  
Db 426 DGCSPVIVKGVVMTYVIGTHRDPISYICADADVFRPERWFEPE-----TRKLGWAYVPFNG 482  
QY 489 GPRICLGKDSAYLQMKMALILFREYSFRLLE---GHPVQVYRMVTLISWAGLKVURV 542  
Db 483 GPRICLGQQAFTASYYTVRLQLQEFGLNLSLDPNAEYPPKLQNTLITLSLFDGADVVRM 539

RESULT 13  
US-08-948-564-16  
Sequence 16, Application US/08948564  
Patent No. 6121512  
GENERAL INFORMATION:  
APPLICANT: Siminszky, Balazs  
APPLICANT: Dewey, Ralph E.  
APPLICANT: Corbin, Frederick T.  
TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and  
TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Virginia C. Bennett  
STREET: PO Box 37428

CITY: Raleigh  
STATE: No. 6121512th Carolina  
COUNTRY: USA  
ZIP: 27627  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,564  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Bennett, Virginia C.  
REGISTRATION NUMBER: 37,092  
REFERENCE/DOCKET NUMBER: 5051-409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-854-1400  
TELEFAX: 919-854-1401  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-948-564-16

Query Match 11.9%; Score 367; DB 3; Length 576;  
Best Local Similarity 24.3%; Pred. No. 1e-31;  
Matches 142; Conservative 96; Mismatches 222; Indels 124; Gaps 24;

QY 16 PF---PFLAGPKYIALLLVLSWILVQWLSRKQKQPR-----S 52  
Db 27 PFTTHFSLSQPKR-----ISSIRCSINTDKKSSRNLLGNANLLTDLISGSGS 78  
QY 53 WPVIGATVEQLRN---YHRMHDWLVGYSRHTVTVDMPFTSYTYIADPVNVEHLKTNF 109  
Db 79 MPAEGAVSLLGRPLFFSLYDN---FLEHGAVVYKLAFGPKAFVWVSDPIVARHILRENA 135  
QY 110 TNPYKGIYRSYMDVLLGDGIFNADGELMRKQRTASFEFASKNLRDFFSAIVREYS--- 166  
Db 136 FSYDKGVL-ADILEPMGKGLIADLDTMKQRRRVIAFAPHNSYL-EAMVKIITTCERT 193  
QY 167 -LKLSGILSQAAGK---VVMQELMYMTLDSICKVGFVEIGTSLPDL-ENSFAQ 220  
Db 194 ILKFNKLEGEVGDGDDPSIELLEAEFSSALDIIIGLVENYDFGVTKSPVIAVYGT 253  
QY 221 AFDAANIIITLRFIDPLWRI---KFFEHVGSALLAQSIKLVDEFTYSVIRRRKAEIVEVR 278  
Db 254 LFEAEH---RSTFYIPYWKIPLARWI-VPRQRKFQDDLKVINTCLDGLIRNAKESROETD 309  
QY 279 ASGQKEK---MKHDIILSRFIEIGAGDDGGGFGDDKSLRDVVLNFVIAGRDTTATLSW 334  
Db 310 VEKLQQRDYLNLKDALSLRLFLV---DMRGADVDDRLRDLMTLMIAGHETFAAVLTW 364  
QY 335 FTHWAMSHPDVAEKLRELCFAFAERAREBGVTILVLCGGADADKAFAFAVAQFAGLLTY 394  
Db 365 AVFLLAQNPSSMKKAQAE-----VDLVLTGRP-----TF 394  
QY 395 DSLGKLVYLHACVTEITRLYPAPQDPKGILEDVLP-----DGTVKRAAGMVTVPY 447  
Db 395 ESLKELQYIRLIVVEALRYLPQPELLIRRSLSKSDVLPFGHKGEKGDYVAIPAGTVFISVY 454  
QY 448 SMGRMEYNWG-PDAASFRPERWI---NED-----GAFRNASP-----FKETAF 486  
Db 455 NLHRSYPFWDRPD---DPEPERFLVQNKNEETEGWAGLDPSRSPGALYXNEVISDFALPFF 512  
QY 487 QAGPRICLGKDSAYLQMKMALILFREYSFRLLEGHVPQVYRMV 530  
Db 513 GGGPRKCVGQDFALMESTVAUTMLLQNPDBE-LKGTPESELVLT 555

Mon Mar 8 10:13:27 2004

RESULT 14  
US-09-302-620B-104  
; Sequence 104, Application US/09302620B  
; Patent No. 6331420  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, C. Ron  
; APPLICANT: Craft, David L.  
; APPLICANT: Birch, Dudley  
; APPLICANT: Eshoo, Mark  
; APPLICANT: Madduri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450  
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA  
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS  
; TITLE OF INVENTION: RELATING THERETO  
; FILE REFERENCE: 1010-16.seg  
; CURRENT APPLICATION NUMBER: US/09/302,620B  
; CURRENT FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 109  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 104  
; LENGTH: 499  
; TYPE: PRT  
; ORGANISM: Candida tropicalis  
US-09-302-620B-104

Query Match 11.5%; Score 353; DB 4; Length 499;  
Best Local Similarity 26.3%; Pred. No. 2.9e-30;  
Matches 127; Conservative 79; Mismatches 182; Indels 94; Gaps 17;  
QY 80 HRVTVDMPFTSYTIADPUNVEHVLKTNFTNPKGIVYRSDVMDVLGGIFNADGELWR 139  
DB 89 NRALSVEPILTK-----DPVNIKAMJSTQDDPSLGLRLHQFAP-LLGKGIFFLDGPEWK 142  
QY 140 KQKRTASFFAS---KNLRDFA--IVFREYSILKSLGILSQAASKAGKAGVMDQMELYMRMTL 194  
DB 143 QSRMLRPQAKORVSHILDLEPHFVLLRKH---IDG-----HNGDYFDIQELYFRFSM 193  
QY 195 DSICKVGFVEICTLSPDLFENSPFAQAFDAANIITLRFIDPLWRKIRFFHVGSEALLAQ 254  
DB 194 DVATGELFGESVGLKDE--DARFLEAFNESQKYLATR-----ATLHE 234  
QY 255 SIKLVDEFTY---SVTRRKAETVEVRASGKOEKVKHDLISRFIELGEAGDDGGGFGDD 310  
DB 235 LYFLCDGFRQYKVKFCSCQVH-----KALDVAPETSSEYFRELVKH---TRDP 286  
QY 311 KSLRDVVLNFVIAGRDITATTLNFWTHMAMSHPDVAEKLRELCAPFAEARBEGVTVL 370  
DB 287 VVLQDQALVLLAGRDTTSLSPATPELARNDHMRKREE-----VIL 331  
QY 371 CGGADDDKAFARVAOAGLITVDSLGKLVYLHACVTETRLYPVAVPQDPKGILEDVVL 430  
DB 332 TMGPSSDE-----ITVAGLKSCRYLKAILNETRLYPSPVRNARFATRNITL 378  
QY 431 PDG-----TKVRAGMVTYVPSYMGMEYNGWPDAAAFRPRERWINEGAFNAPFFK 482  
DB 379 PRGGPDGSPFILLRKQPVGYFCATHLNEKVGNDSHVFRPRERWAALEG---XSLGWS 435  
QY 483 FTAFQAGPRICLGKDSAYLQMMALAILFRFYS---FLLLEGHPVQVYRMMTILSMAGLK 539  
DB 436 YLPFNGGPRSCLGQCFALLESAYVLARLTQCYTTIQLRTTETPPPKLVHLT-MSLLNGVY 494  
QY 540 VR 541  
DB 495 IR 496

RESULT 15  
US-09-976-594-533

; Sequence 533, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PEARL Program  
; SEQ ID NO 533  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 6825202CD1  
US-09-976-594-533

Query Match 11.5%; Score 353; DB 4; Length 524;  
Best Local Similarity 23.8%; Pred. No. 3.2e-30;  
Matches 138; Conservative 97; Mismatches 226; Indels 118; Gaps 19;  
QY 9 LTPATPSPPFPLAGPHKXIALLLVLSWILYQMSLRKQKGRSPWPVIGATVQQLRNYHR 68  
DB 11 LGPVAASP-----WLLLLLVGGSWLLAR-----VLAWTYTFYDNCRR 47  
QY 69 M-----HDLVGYL-----SRHRTVTVDMPFTSY-----TYIADPVNV 101  
DB 48 LQCFQPPQKQNFQWGHQGLVPTTEGKMTLTQLVITYQGFKLWLGPTFPLLLCHPDII 107  
QY 102 EHVLTNFTNPKGIVYRSDVMDVLGGIFNADGELMRKQRTAFSEFASKNLRDFSALV 161  
DB 108 RPIITSASAAPAAPKDMIFGFLKPLWGLDGLLLSGGKWSRHRRLMLTAPAFENILKPKMKIF 167  
QY 162 FREYSLKSLGILSQAASKAGKAGVMDQMELYMRMTLDSICKVGFVEIGT-----LSPDLP 214  
DB 168 NKSVMIMDKWQRLASGSAKLDMPFHSMTMLDLSLQKCVFSEFNSCQKPEYIAALIE 227  
QY 215 ENSEAFQAPDAANIIT--LRFIDP--LWRIKRFHVHVGSEALLAQSIKLVDEFYTSVIRRK 271  
DB 228 LSFAVEKRNQQLLHHTDFLYITPDGQFRACH-----LVHDFTDVAVIQER 275  
QY 272 ABIVEVRASGKOEKVKHDLISRFIELGE---AGDDGGFGDDKSLRDVVLNMFVIAGRT 327  
DB 276 RTL---PTQGIIDFLKNKAKSKTLDIFDVLVLLSKDDEKELSDDEDIRABADTFMEGHDT 332  
QY 328 TATTLNFWTHMAMSHPDVAEKLRELCAPFAEARBEGVTVLVLCGSDADDDKAFARVAQ 387  
DB 333 TASGLSWILYHLAKHPEYQECRQEVQCELLKORBPTE----- 369  
QY 388 FAGLLFYDSIGKLVYLHACVTETRLYPVAVPQDPKGILEDVVLPGTKVRAAGMVTYVY 447  
DB 370 -----IEWDDLAQLPFTIMIKESLRLLHPVPVISRCTQDFVLPDG-RVPIKXIVCLI-- 422  
QY 448 SMGRMEYN---MGPDAAFRPRERWINEGAFNAPSPKFTAFQAGPRICLGKDSAYLQWK 504  
DB 423 NIIGIHYNPVW-PDPEVDYDPPFRDQEN--IKERSPLAFIPFSAGPRNCIGQAFAMAEMK 479  
QY 505 MALAILFRFYSFLLLEGHPVQVYRMMTILSMH-GLKVRV 542  
DB 480 VVLAL--TLHFRILPTHTPEPRKRPPELLILRAEGGLIWRV 516

Search completed: March 2, 2004, 15:30:19  
Job time : 24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:21:34 ; Search time 60 Seconds  
(without alignments)  
2768.966 Million cell updates/sec

Title: US-10-021-657-2

Perfect score: 3079

Sequence: 1 EFGTREAHLPATPPFPPL.....ICMHACKGRWVSLVAVLKP 588

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3079	100.0	588	5	ABB08037
2	976	31.7	541	2	AA01648
3	976	31.7	541	2	AA01647
4	971.5	31.6	537	3	AA01647
5	970.5	31.5	530	3	AA01647
6	969.5	31.5	523	3	AA01647
7	961.5	31.2	522	3	AA01647
8	961.5	31.2	527	3	AA01647
9	911.5	29.6	489	3	AA01647
10	894	29.0	475	3	AA01647
11	886	28.8	475	3	AA01647
12	884	28.7	465	3	AA01647
13	847	27.5	508	3	AA01647
14	847	27.5	510	3	AA01647
15	847	27.5	528	3	AA01647
16	841	27.3	508	3	AA01647
17	841	27.3	510	3	AA01647
18	841	27.3	527	3	AA01647
19	840	27.3	512	2	AA01647
20	832	27.0	511	2	AA01647
21	831.5	27.0	510	2	AA01647
22	818.5	26.6	513	2	AA01647
23	802.5	26.1	511	2	AA01647
24	785.5	25.5	514	2	AA01647
25	780	25.3	519	5	ABB90378

26	773.5	25.1	477	3	AA023271	Arabidops
27	770.5	25.0	448	3	AA023272	Arabidops
28	770	25.0	158	5	ABB08038	Sorghum S
29	760.5	24.7	464	6	ADA48304	Ada48304 Rice proc
30	747.5	24.3	438	3	AA023273	Arabidops
31	744	24.2	516	3	AA023273	Arabidops
32	738.5	24.0	510	3	AA023273	Arabidops
33	730	23.7	479	3	AA023273	Arabidops
34	728	23.6	475	3	AA023273	Arabidops
35	727.5	23.6	506	3	AA023273	Arabidops
36	726.5	23.6	437	3	AA023273	Arabidops
37	721.5	23.4	495	3	AA023273	Arabidops
38	721.5	23.4	497	3	AA023273	Arabidops
39	721.5	23.4	507	3	AA023273	Arabidops
40	720	23.4	441	3	AA023273	Arabidops
41	693.5	22.5	464	3	AA023273	Arabidops
42	693.5	22.5	471	3	AA023273	Arabidops
43	687	22.3	428	3	AA023273	Arabidops
44	685	22.2	463	3	AA023273	Arabidops
45	685	22.2	467	3	AA023273	Arabidops

## ALIGNMENTS

RESULT 1  
ABB08037  
ID ABB08037 standard; protein; 588 AA.  
XX  
AC ABB08037;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Maize SBMu200 protein.  
XX  
KW SBMu200; fertility; maize; hybrid seed; sterility; plant; transgenic.  
XX  
OS Zea mays.  
XX  
FH Key Location/Qualifiers  
FT Region 1..546  
FT /note= "N-terminal portion of SBMu200"  
FT Region 546..547  
FT /note= "point of fusion between the N- and C- terminal parts of SBMu200, separated by a stop codon in the polynucleotide sequence"  
FT Region 547..588  
FT /note= "C-terminal portion of SBMu200"  
XX  
WO200226789-A2.  
XX  
PD 04-APR-2002.  
XX  
PF 25-SEP-2001; 2001WO-US029886.  
XX  
PR 26-SEP-2000; 2000US-00670153.  
XX  
PA (PTON-) PIONEER HI-BRED INT INC.  
XX  
PI Albertsen MC, Fox T, Huffman G, Trinnell M;  
XX  
DR WPI; 2002-471258/50.  
XX  
N-PSDB; ABL60611.  
XX  
Novel nucleic acid sequence comprising SBMu200 gene, useful for producing transgenic plants and mediating male fertility in plants.  
XX  
Claim 2; Fig 4; 60pp; English.  
XX  
The invention relates to a SBMu200 gene, useful for impacting male fertility in a plant e.g. maize. An isolated nucleotide sequence comprising the SBMu200 gene is useful for producing hybrid seeds and for providing heritable externally controllable male sterility in a plant.

CC The present sequence represents the maize SBW1200 protein

XX  
SQ Sequence 588 AA;  
Query Match 100.0%; Score 3079; DB 5; Length 588;  
Best Local Similarity 100.0%; Pred. No. 1.8e-303;  
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EFGTREAHTPATPSPPFFLAGPHKVIALLVLSWILVQMSLRKQKQKPRSPWVIGATV 60  
DB 1 EFGTREAHTPATPSPPFFLAGPHKVIALLVLSWILVQMSLRKQKQKPRSPWVIGATV 60  
QY 61 EQLRNHRHEDWLVGLVLSHRHTVDMPTSTYTIADPNVVEHLKTNFTNPKGIVYES 120  
DB 61 EQLRNHRHEDWLVGLVLSHRHTVDMPTSTYTIADPNVVEHLKTNFTNPKGIVYES 120  
QY 121 YMDVLGDI FNADEGLWKRKORTASFEFASKNLRDPSAIVREYSKLGLSLQASKAG 180  
DB 121 YMDVLGDI FNADEGLWKRKORTASFEFASKNLRDPSAIVREYSKLGLSLQASKAG 180  
QY 181 KVVDMOELVMTLDSICKVGFVEIGTLSPDLSPNSFAQAFDAANIITLRFIDELWRI 240  
DB 181 KVVDMOELVMTLDSICKVGFVEIGTLSPDLSPNSFAQAFDAANIITLRFIDELWRI 240  
QY 241 KRFFHYGSEALLAQSIKLVDFTYVIRRRKABIEVVRASGKQKMKHDLILSRFIELGEA 300  
DB 241 KRFFHYGSEALLAQSIKLVDFTYVIRRRKABIEVVRASGKQKMKHDLILSRFIELGEA 300  
QY 301 GDDGGGGDDKSLRDVVLNFEVIAGRDTTATLTSWETHMAMSHPDVAEKLRLRELCAFEAER 360  
DB 301 GDDGGGGDDKSLRDVVLNFEVIAGRDTTATLTSWETHMAMSHPDVAEKLRLRELCAFEAER 360  
QY 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVAPQD 420  
DB 361 AREEGVTLVLCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVAPQD 420  
QY 421 PKGILEDDVLPDGTQVRAGGWTVPYSGRMEYNWGPDAASPRPERWINEDEGAFRNASP 480  
DB 421 PKGILEDDVLPDGTQVRAGGWTVPYSGRMEYNWGPDAASPRPERWINEDEGAFRNASP 480  
QY 481 FKFTAFQAGPRICLGKDSAYLQMKWALAILFRFYSFRLLEGHVPQYRMNTILSMAHGLKV 540  
DB 481 FKFTAFQAGPRICLGKDSAYLQMKWALAILFRFYSFRLLEGHVPQYRMNTILSMAHGLKV 540  
QY 541 RVSRVAVCHGDLMDIVPLNPRQITLVLIQICMCHACKGRWVSLVAVLKP 588  
DB 541 RVSRVAVCHGDLMDIVPLNPRQITLVLIQICMCHACKGRWVSLVAVLKP 588

RESULT 2  
AA01648  
ID AA01648 standard; protein; 541 AA.

XX  
AC AA01648;  
XX  
DT 23-JUN-1999 (first entry)  
XX  
DE Cytochrome P450 encoded by a codon modified cDNA.  
XX  
KW Cytochrome P450; protein expression; yeast; mutant.  
XX  
OS Synthetic.  
OS Triticum aestivum.  
XX  
PN FR2768748-A1.  
XX  
PD 26-MAR-1999.  
XX  
PF 24-SEP-1997; 97FR-00012094.  
XX  
PR 24-SEP-1997; 97FR-00012094.  
XX  
PA (RHON ) RHONE-FOULENC AGROCHIMIE.

XX Batard Y, Schalk M, Durst F, Werck RD;

XX WPI; 1999-217499/19.  
XX N-PSDB; AAX26878.

XX DNA for expression in yeasts - with codon changes based on yeast codon usage.

XX Example 4; Page 24-26; 31pp; French.

XX The present sequence represents a cytochrome P450, encoded for by a cDNA which was created by altering the codons of AAX26865 to exemplify the invention. The specification describes a DNA sequence that encodes a protein of interest and contains regions with a high content of codons poorly suited to yeasts, where a sufficient number of these codons are replaced by corresponding codons suited to yeasts in these regions. Yeasts transformed with vectors such DNA sequences can be cultured to produce the protein of interest (especially a plant protein) or, when the protein is an enzyme (such as cytochrome P450), can be cultured in the presence of a substrate for the enzyme to produce a substrate conversion product

XX Sequence 541 AA;

Query Match 31.7%; Score 976; DB 2; Length 541;

Best Local Similarity 39.9%; Pred. No. 1.2e-89;  
Matches 222; Conservative 82; Mismatches 190; Indels 62; Gaps 13;

QY 27 IALLVLSWILVQMSLRKQKQKPRSPWVIGATVQELRNHRHEDWLVGLYLSR---HR 81  
DB 10 VSAAVAAWAF---WRMSRGLGRFVWVLSGLPLGVQHAEDMHEWIAGNLRRAGGTQ 65  
QY 82 TVTVDMPFTS---YTYIADPNVVEHLKTNFTNPKGIVRYSYMDVLLGGINADGE 136  
DB 66 TCIFAVGVARRGGLVTTCDPNLEHLVKARFDNYPKGFPHVGVFRDVLGSGIFNSDGD 125  
QY 137 LWRKORKTASFASFASKNLRD-FSAIVREYSKLGLSLQASKAGKQKQKPRSPWVIGATV 195  
DB 126 TWLAQKRTAALBETTTTLTATMSRWVRSIHGRLPLILADAAGKQKQKQVLDLALLLTFD 185  
QY 196 SICKVGFVGEIGTSLSPDLSPNSFAQAFDAANIITLRFIDP--LWPKRFFHVGSALLA 253  
DB 186 NICGLAFKDPETLAQGLPENEFASAFDRATEATLNRFFPEFLWCKKWLGLMETTLT 245  
QY 254 QSTKLVDFTYVIRRRKABIEVVRASGKQKMKHDLILSRFIELGAGDDGGFGDDKSL 313  
DB 246 SSMAHVDQVLAIVIKRKLALAGNGKCDTAAATHDILLSRFMRKGSYSDE-----SL 297  
QY 314 RDVVLNFEVIAGRDTTATLTSWETHMAMSHPDVAEKLRLRELCAFEARAREEGVTLVLCGG 373  
DB 298 QHVALNFIAGRDITSSVALSWFFVLVSTHFAVERKIVRELCS-----VLAAS 344  
QY 374 ADADKAFARVAQFAGLLTYDSLGKLVYLHACVTETLRLYPVAPQDPKGILEDDVLPD 433  
DB 345 RGAHPAL-----WLAEPFTFEELDRILVYLKAALSETLRLYSPVPEDSKHVADYLPD 399  
QY 434 TKVRAGGWTVPYSGRMEYNWGPDAASPRPERWINEDEGA-FRNASPFKFTAFQAGPRI 492  
DB 400 TFVPAGSSVTYISYAGRMKGVWGEDCLEFRPRRWLSADGTFKEQHDYSKFFVAFNAGPRV 459  
QY 493 CLGKDSAYLQMKWALAILFRFYSFRLLEGHVPQYRMNTILSMAHGLKQKQKPRSPWVIGATV 551  
DB 460 CLGKDLAYLQMKWALAILFRFYSFRLLEGHVPQYRMNTILSMAHGLKQKQKPRSPWVIGATV 551  
QY 552 DMDIVPLNPRQITLV 567  
DB 510 -----RPRDLAPVL 518

RESULT 3  
AA01647  
ID AA01647 standard; protein; 541 AA.

XX	AAV01647;	Db	345	RGADPAL-----WLAEPFTFEELDRVLKALSETLRLYSPVEDSKHVVDYDLPDG	399
AC					
XX	23-JUN-1999 (first entry)	QY	434	TKVRAGGMVTVPIYSKGRMEYNWGDPAASFRPERWINEGGA-FRNASPKFTAFQAGPRI	492
XX					
DE	Cytochrome P450 designated CYP86A5.	Db	400	TFVPAGSSVTYSIYSAGRMKGVMGEDCCLFPRERWLSADGTFKFEQHDSYKFFVAFNAGPRV	459
XX					
KW	Cytochrome P450; protein expression; yeast.	QY	493	CLGKDSAYLQMK-MALAJLFRYSFRLLEGHVPQVYRMWMTILSMAHGLKVRVSRVAVCHGDL	551
XX					
OS	Triticum aestivum.	Db	460	CLGKDLAYLQMKNIAGSVLLR-HRLTVAPGHRVQKMSLTFLMKGLRMEV-----	509
XX					
XX	FR2768748-A1.	QY	552	DMDIVELNPRQITLWL	567
XX					
XX	26-MAR-1999.	Db	510	-----RPRDLAPVL	518
XX					
XX	24-SEP-1997; 97PR-00012094.				
XX					
XX	24-SEP-1997; 97PR-00012094.				
XX	(RHON ) RHONE-POULENC AGROCHIMIE.				
XX					
PI	Bataud Y, Schalk M, Durst F, Werck RD;				
XX					
DR	WPI; 1999-217499/19.				
DR	N-PSDB; AAX26874.				
XX	DNA for expression in yeasts - with codon changes based on yeast codon				
PT	usage.				
XX					
PS	Example 4; Page 21-23; 31pp; French.				
XX					
CC	The present sequence represents a cytochrome P450, and was used to				
CC	exemplify the invention. The specification describes a DNA sequence that				
CC	encodes a protein of interest and contains regions with a high content of				
CC	codons poorly suited to yeasts, where a sufficient number of these codons				
CC	are replaced by corresponding codons suited to yeasts in these regions.				
CC	Yeasts transformed with vectors such DNA sequences can be cultured to				
CC	produce the protein of interest (especially a plant protein) or, when the				
CC	protein is an enzyme (such as cytochrome P450), can be cultured in the				
CC	presence of a substrate for the enzyme to produce a substrate conversion				
CC	product				
XX					
XX	Sequence 541 AA;				
XX					
XX	Query Match				
XX	Best Local Similarity 31.7%; Score 976; DB 2; Length 541;				
XX	Matches 222; Conservative 82; Mismatches 190; Indels 62; Gaps 13;				
QY	27 IALLVLSWILVQWMSL-RKQKGRSPWPVIGATVEQLRNVHRMHDWLVGYLSR-----HR	81			
Db	10 VSAVAAYMAWF-----WRMSRGLRPRVWPVLGSLGVLQVHAEDMHEWTAGNLRRAAGGTIQ	65			
QY	82 TVTVDMPTFS-----YTYIADPNVVEHVLKNTFTYKPGIVYRSYMDVLLGDGFIENADGE	136			
Db	66 TCIFAVGVARGGLVTVTCDPNLEHVLKARFDNYPKGFPHGVRFDLLGDGFIENSDGD	125			
QY	137 LMRKORTASFEFASKNLRD-FSAIVFREYSKLSGLISQAKGKQVMDMQLYRMWILD	195			
Db	126 TWLAQRKTAALFTTTLTANTSMRWVSRSIHGELLPIILADAAKGAQVDQLLLRLTFD	185			
QY	196 STCKVGFGEIOTLSPDLPENSAQAFDAANIITLRFIDP--LWRIKFFPHVSGEALLA	253			
Db	186 NICGLAFGKDPETLAQGLPENEFASAFDRATEATLNRFFPFBLMRCKKWLGLMETTTT	245			
QY	254 QSIKLVDETYSVIRERKAEIIEVRASGQKMKHDIILSRFTELGRAGDGGGFGDKSL	313			
Db	246 SSMHVDQYLAIVIKKKLELAAGNCKDCTAATHDLLSRFKGKSYSDS-----SL	297			
QY	314 RDVILNFVIAGRTTATILSWFTTHMAMSHPDVAEKLRLREICAFAPRAREEGVTILVLCGG	373			
Db	298 QHVALNFILAGRTSSVALSWFFFLVSTHFAVERKIVRELCS-----VLAAS	344			
QY	374 ADADKAFARVAQFAGLLTYSGLKLVYLHACVTETLRLYPAVPQDPKGILEDVLPDG	433			

us-10-021-657-2.rag

Mon Mar 8 10:13:27 2004

PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 21-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140891P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142055P.  
PR 02-JUL-1999; 99US-0142390P.  
PR 06-JUL-1999; 99US-0142803P.  
PR 08-JUL-1999; 99US-0142920P.  
PR 09-JUL-1999; 99US-0142977P.  
PR 12-JUL-1999; 99US-0143542P.  
PR 13-JUL-1999; 99US-0143624P.  
PR 14-JUL-1999; 99US-0144005P.  
PR 15-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 17-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 23-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 25-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.



Mon Mar 8 10:13:27 2004

us-10-021-657-2.rag

Query Match	31.6%; Score 971.5; DB 3; Length 537;	
Best Local Similarity	40.9%; Pred. No. 3.3e-89;	
Matches 217; Conservative 87; Mismatches 182; Indels 45; Gaps 12;		
QY	27 IALLVVLVSLVQSLKQKQPSWPIVIGATVEOLRNVRHMDLVGLV---SRHET 82	
Db	7 IMLISAITAYFLLMTISCLKQPRVWPPIIGSLPGLIENCERHMDVSNLRACSTYGT 66	
QY	83 VTVDMEFTS-----YTIADPPNVNVEHLKTNFTNPKGIVYRSYMDVLLGDGIFNADGEL 137	
Db	67 CITCAIFLAKKQGLVTVTCDPRLNLEHLKORFNDYKPGTQWQAVEHDLGQGFNSDGD 126	
QY	138 WRKQRTASFEFASKNLRDPSA-IVPREYLSKUSGILSQASKAGKVVDMQELVMTLDS 196	
Db	127 WLFQRTAALEFTRTLQAMARWNRRAIKRLPFIENARLGSBPDLDQLLRLTFDN 186	
QY	197 ICKVGVGEIGTSLDPENSPQAQDAANIITLRFIDP--LWRIKRFPHVGEALLAQ 254	
Db	187 ICGLTGKDPRTCAPGLPVNTFAVDRAEASLQRFILPEILWKFRLGLEVSLTR 246	
QY	255 SIKLVDEFTYSVIRRRKAEIVEVRASQKQKMKHDILSRFIELGEAGDDGGFGDDKSLR 314	
Db	247 SLVQVDNYLSEIITRKEENMTQHNGKHH---DGLLSRFIKKES-----YSDETQ 296	
QY	315 DVVLNFIAGRDATTATLSWFTWAMSHDPVAEKLRELCAFEARAREEGVTIVLCGGA 374	
Db	297 RVALNFTLAGRDTSSVALSWFFMLITQHPAIEDKILREICTVILVE-TRGDDVALW----- 350	
QY	375 DADDKFAARVAQFAGLLTYDSLGLVYLHACVTETLRLYPAPQDPKGILEDVLPDGT 434	
Db	351 -TDEP-----LSCELDRLVFLKAAISETLRLYPSVPEDSKRAVKDDVLPDGT 397	
QY	435 KVRAGGMVTVPYSGRMEXNWGPDAAASFRPERWINE--GAFENASPPKFTAFQAGPRI 492	
Db	398 FVPAGSSITYSIYAGRMKSTWGEDCLEFKEPERWISQDGRFINHDPFKFVAFNAGPRI 457	
QY	493 CLGKDSAYLQWK-MALAILFRFYSFRLLGHPVQYRMVMTILSMAGHLKVRV 542	
Db	458 CLGKDLAYLQWKSIASAVLLR-HRLTVVTGKHVEQKVSLLTLMKYGLLVNV 507	
RESULT 5		
AAAG36728		
ID	AAAG36728 standard; protein; 530 AA.	
XX	AAAG36728;	
AC		
XX		
DT	18-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 45051.	
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
OS	Arabidopsis thaliana.	
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-00301439.	
XX		
PR	25-FEB-1999; 99US-0121825P.	
PR	05-MAR-1999; 99US-0123180P.	
PR	09-MAR-1999; 99US-0123548P.	
PR	23-MAR-1999; 99US-0125788P.	
PR	25-MAR-1999; 99US-0126264P.	
PR	29-MAR-1999; 99US-0126785P.	
PR	01-APR-1999; 99US-0127462P.	
PR	06-APR-1999; 99US-0128234P.	
PR	08-APR-1999; 99US-0128714P.	
PR	16-APR-1999; 99US-0129845P.	
PR	19-APR-1999; 99US-0130077P.	
PR	21-APR-1999; 99US-0130449P.	
PR	23-APR-1999; 99US-0130510P.	
PR	23-APR-1999; 99US-0130891P.	
PR	28-APR-1999; 99US-0131449P.	
PR	30-APR-1999; 99US-0132048P.	
PR	30-APR-1999; 99US-0132407P.	
PR	04-MAY-1999; 99US-0132484P.	
PR	05-MAY-1999; 99US-0132485P.	
PR	06-MAY-1999; 99US-0132486P.	
PR	06-MAY-1999; 99US-0132487P.	
PR	07-MAY-1999; 99US-0132863P.	
PR	11-MAY-1999; 99US-0134256P.	
PR	14-MAY-1999; 99US-0134218P.	
PR	14-MAY-1999; 99US-0134219P.	
PR	14-MAY-1999; 99US-0134221P.	
PR	14-MAY-1999; 99US-0134370P.	
PR	18-MAY-1999; 99US-0134768P.	
PR	19-MAY-1999; 99US-0134941P.	
PR	20-MAY-1999; 99US-0135124P.	
PR	21-MAY-1999; 99US-0135353P.	
PR	24-MAY-1999; 99US-0135629P.	
PR	25-MAY-1999; 99US-0136021P.	
PR	27-MAY-1999; 99US-0136392P.	
PR	28-MAY-1999; 99US-0136782P.	
PR	01-JUN-1999; 99US-0137222P.	
PR	03-JUN-1999; 99US-0137528P.	
PR	04-JUN-1999; 99US-0137502P.	
PR	07-JUN-1999; 99US-0137724P.	
PR	08-JUN-1999; 99US-0138094P.	
PR	10-JUN-1999; 99US-0138540P.	
PR	10-JUN-1999; 99US-0138847P.	
PR	14-JUN-1999; 99US-0139119P.	
PR	16-JUN-1999; 99US-0139452P.	
PR	16-JUN-1999; 99US-0139453P.	
PR	17-JUN-1999; 99US-0139492P.	
PR	18-JUN-1999; 99US-0139454P.	
PR	18-JUN-1999; 99US-0139455P.	
PR	18-JUN-1999; 99US-0139456P.	
PR	18-JUN-1999; 99US-0139457P.	
PR	18-JUN-1999; 99US-0139458P.	
PR	18-JUN-1999; 99US-0139459P.	
PR	18-JUN-1999; 99US-0139460P.	
PR	18-JUN-1999; 99US-0139461P.	
PR	18-JUN-1999; 99US-0139462P.	
PR	18-JUN-1999; 99US-0139463P.	
PR	18-JUN-1999; 99US-0139750P.	
PR	18-JUN-1999; 99US-0139763P.	
PR	21-JUN-1999; 99US-0139817P.	
PR	22-JUN-1999; 99US-0139899P.	
PR	23-JUN-1999; 99US-0140353P.	
PR	23-JUN-1999; 99US-0140354P.	
PR	24-JUN-1999; 99US-0140695P.	
PR	28-JUN-1999; 99US-0140823P.	
PR	29-JUN-1999; 99US-0140991P.	
PR	30-JUN-1999; 99US-0141287P.	
PR	01-JUL-1999; 99US-0141842P.	
PR	01-JUL-1999; 99US-0142154P.	
PR	02-JUL-1999; 99US-0142055P.	
PR	06-JUL-1999; 99US-0142390P.	
PR	08-JUL-1999; 99US-0142803P.	
PR	09-JUL-1999; 99US-0142920P.	
PR	12-JUL-1999; 99US-0142977P.	
PR	13-JUL-1999; 99US-0143542P.	
PR	14-JUL-1999; 99US-0143624P.	
PR	15-JUL-1999; 99US-0144005P.	
PR	16-JUL-1999; 99US-0144085P.	
PR	16-JUL-1999; 99US-0144086P.	
PR	19-JUL-1999; 99US-0144325P.	
PR	19-JUL-1999; 99US-0144331P.	
PR	19-JUL-1999; 99US-0144332P.	
PR	19-JUL-1999; 99US-0144333P.	
PR	19-JUL-1999; 99US-0144334P.	

```

19-JUL-1999; 99US-0144335P.
20-JUL-1999; 99US-0144352P.
20-JUL-1999; 99US-0144632P.
20-JUL-1999; 99US-0144884P.
21-JUL-1999; 99US-0144814P.
21-JUL-1999; 99US-0145086P.
21-JUL-1999; 99US-0145088P.
22-JUL-1999; 99US-0145085P.
22-JUL-1999; 99US-0145087P.
22-JUL-1999; 99US-0145089P.
22-JUL-1999; 99US-0145192P.
23-JUL-1999; 99US-0145145P.
23-JUL-1999; 99US-0145218P.
23-JUL-1999; 99US-0145224P.
26-JUL-1999; 99US-0145276P.
27-JUL-1999; 99US-0145913P.
27-JUL-1999; 99US-0145918P.
27-JUL-1999; 99US-0145919P.
28-JUL-1999; 99US-0145951P.
02-AUG-1999; 99US-0146386P.
02-AUG-1999; 99US-0146388P.
02-AUG-1999; 99US-0146389P.
03-AUG-1999; 99US-0147038P.
04-AUG-1999; 99US-0147204P.
04-AUG-1999; 99US-0147302P.
05-AUG-1999; 99US-0147192P.
05-AUG-1999; 99US-0147260P.
06-AUG-1999; 99US-0147303P.
06-AUG-1999; 99US-0147416P.
08-AUG-1999; 99US-0147493P.
08-AUG-1999; 99US-0147935P.
10-AUG-1999; 99US-0148171P.
11-AUG-1999; 99US-0148319P.
12-AUG-1999; 99US-0148341P.
13-AUG-1999; 99US-0148565P.
13-AUG-1999; 99US-0148684P.
16-AUG-1999; 99US-0149368P.
17-AUG-1999; 99US-0149175P.
18-AUG-1999; 99US-0149426P.
20-AUG-1999; 99US-0149722P.
20-AUG-1999; 99US-0149723P.
20-AUG-1999; 99US-0149929P.
23-AUG-1999; 99US-0149902P.
23-AUG-1999; 99US-0149930P.
25-AUG-1999; 99US-0150566P.
26-AUG-1999; 99US-0150884P.
27-AUG-1999; 99US-0151065P.
27-AUG-1999; 99US-0151066P.
27-AUG-1999; 99US-0151080P.
30-AUG-1999; 99US-0151303P.
31-AUG-1999; 99US-0151438P.
01-SEP-1999; 99US-0151930P.
07-SEP-1999; 99US-0152363P.
10-SEP-1999; 99US-0153070P.
13-SEP-1999; 99US-0153758P.
15-SEP-1999; 99US-0154018P.
16-SEP-1999; 99US-0154039P.
20-SEP-1999; 99US-0154779P.
22-SEP-1999; 99US-0155139P.
23-SEP-1999; 99US-0155486P.
24-SEP-1999; 99US-0155659P.
28-SEP-1999; 99US-0156458P.
29-SEP-1999; 99US-0156596P.
04-OCT-1999; 99US-0157117P.
05-OCT-1999; 99US-0157533P.
06-OCT-1999; 99US-0157865P.
07-OCT-1999; 99US-0158023P.
08-OCT-1999; 99US-0158232P.
12-OCT-1999; 99US-0158369P.
13-OCT-1999; 99US-0159293P.
13-OCT-1999; 99US-0159294P.
13-OCT-1999; 99US-0159295P.
14-OCT-1999; 99US-0159323P.
99US-0159330P.
99US-0159331P.
99US-0159637P.
99US-0159638P.
99US-0159584P.
99US-0160741P.
99US-0160767P.
99US-0160768P.
99US-0160770P.
99US-0160814P.
99US-0160815P.
99US-0160980P.
99US-0160981P.
99US-0160989P.
99US-0161404P.
99US-0161405P.
99US-0161406P.
99US-0161359P.
99US-0161360P.
99US-0161361P.
99US-0161920P.
99US-0161992P.
99US-0161993P.
99US-0162142P.
Query Match 31.5%; Score 970.5; DB 3; Length 530;
Best Local Similarity 41.0%; Pred. No. 4e-89;
Matches 217; Conservative 86; Mismatches 181; Indels 45; Gaps 12;
QY 29 LLLVLSLWLVQSLRKQKSPRSEVIGATVEQLRNHYRHMHWLVGYL-----SHRVTWT 84
DB 2 ILSAITAVFLWLTFTISRCCLKGPRVWPIGLSLPLGIENCERMHWISDNLRACSGYTQTCI 61
QY 85 VDMPTFS-----YTYIADPVNTEHVLTNFTNYPKGIYVRSYMDVLGLDGIENAGELWR 139
DB 62 CAIPELAKKQGLVTTCDDPRNLEHLKRNFDNPKGPTWQAVFHLGLLGGIENSGDWTWL 121
QY 140 KOKTASFEFASKNLRFDSA-IVFREYSLSKLSGLSQASKAGKGVMDQELYNRMWTLDSIC 198
DB 122 FORKTAALFTTTLRQAVARNWNRKURFLPILENARLSEPIDLQDLALLRFLPDNIC 181
QY 199 KVGFGVEIGTSLDLPENSFAQAFDAANIITLRFIDP--LWRIKRFHVHVGSEALLAQST 256
DB 182 GLTFGKDPRTCAPCLPVNTFAVAFDEATASLQRFILPEILWKEKWLRLGLEVSLTRSL 241
QY 257 KLVDFTYSVIRRRKAEIIVEVRASQKQKMKHDIILSRFIELGHAGDDGGFGDDKSLRDV 316
DB 242 VQVDNYLSBIITTRKEENMTQHNNGKH---DOLLSRFICKES-----YSDETLQRV 291
QY 317 VLNFIAGRDTTATLTSWFTMHAMSHPDVAEKLRLRELCAPAEARAREEGVTVLVLCGADA 376
DB 292 ALNFIAGRDTSVALSWFFWLLITQHPAIEDKILREICTVLVE-TRGDDVALW-----T 344
QY 377 DDKFAARVAQAQAGLTYDSLGLKLVYLHACVTELTRLYPVAVPODPKGILEDDVLPDGTKV 436
DB 345 DEP-----LSCELDRLVFLKAALSETLRLYSPVSPEDSKRAVKDDVLPDGTFFV 392
QY 437 RAGGMVTVYPYSGRMEYVWGPDAASFRPERWLNED--GAFRNASPKFTAFQAGPRICL 494
DB 393 PAGSSITYSYISAGRMKSTWGECLEFKPERWISQSDGGRFINHDFFKFAVFNAGPRICL 452
QY 495 GKDSAYLOMK-MALAILFRFYSFRLLEGHVPQVYRMMTILSMHAGLKVYV 542
DB 453 GKDLAYLOMKSIASAVLLR-HRLTVTVTGKVEQNSLTLFMKYGLLVNV 500
RESULT 6
AAG54203
ID AAG54203 standard; protein; 523 AA.
XX AAG54203;
XX AC
XX AC
DT 18-OCT-2000 (first entry)

```

Mon Mar 8 10:13:27 2004

us-10-021-657-2.rag

XX Arabidopsis thaliana protein fragment SEQ ID NO: 69087.  
DE Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
KW  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 25-APR-1999; 99US-0130891P.  
XX 28-APR-1999; 99US-0131449P.  
XX 30-APR-1999; 99US-0132048P.  
XX 30-APR-1999; 99US-0132407P.  
XX 04-MAY-1999; 99US-0132484P.  
XX 05-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 06-MAY-1999; 99US-0132487P.  
XX 07-MAY-1999; 99US-0132863P.  
XX 11-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 04-JUN-1999; 99US-0137502P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139492P.  
XX 18-JUN-1999; 99US-0139454P.  
XX 18-JUN-1999; 99US-0139455P.  
XX 18-JUN-1999; 99US-0139456P.  
XX 18-JUN-1999; 99US-0139457P.  
XX 18-JUN-1999; 99US-0139458P.  
XX 18-JUN-1999; 99US-0139459P.  
XX 18-JUN-1999; 99US-0139460P.  
XX 18-JUN-1999; 99US-0139461P.  
XX 18-JUN-1999; 99US-0139462P.  
XX 18-JUN-1999; 99US-0139463P.  
XX 18-JUN-1999; 99US-0139750P.  
XX 18-JUN-1999; 99US-0139763P.

PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140363P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142053P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 22-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147316P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.

PR	27-AUG-1999;	99US-0151080P.	QY	314	RDVVLNFVIAGRDTTATTLSWFTMHAMSHDVAEKLRRELCAFEAEARARBEQVTLVLCGG	373
PR	30-AUG-1999;	99US-0151303P.	Db	296	KYVALNFILAGRDTSSVAMSWFFWLVSINPRVEKIINEIC-----TILI---	340
PR	31-AUG-1999;	99US-0151438P.				
PR	01-SEP-1999;	99US-0151930P.	QY	374	ADADDKAFARVAQFAG-LTYDSLGLKLYLHACVTTTLRLYPAPVDPKGILEDVLPD	432
PR	07-SEP-1999;	99US-0152363P.	Db	341	-----KTRDTNVSXWTDPLTDFEIDQLVYLKALSETLRLYSPVEDSKFVVANDVLPD	395
PR	10-SEP-1999;	99US-0153070P.				
PR	13-SEP-1999;	99US-0153758P.	QY	433	GTKVRAGMVTYVPYSNGRMEYNWGPDAASFRPERWINEGAFRNAGFEKFTAFQAGPRI	492
PR	15-SEP-1999;	99US-0154018P.	Db	396	GTFFVSGSNVTYSTYSVGRMKFIWGEDCLCFKPERWL-EESRDEKCKQYKFVAENAGPRI	454
PR	16-SEP-1999;	99US-0154039P.				
PR	20-SEP-1999;	99US-0154779P.	QY	493	CLGKDSAYLOWKMLAILFRFYSFRLLEGHPVOYRMMTILSMAHGLKVRYSR	544
PR	22-SEP-1999;	99US-0155139P.	Db	455	CLGKDLAYLQMKSIATASILLRHRITVAPGRHVEQKMSLTLFMKFLKMDVHK	506
PR	23-SEP-1999;	99US-0155488P.				
PR	24-SEP-1999;	99US-0155659P.				
PR	28-SEP-1999;	99US-0156458P.				
PR	29-SEP-1999;	99US-0156596P.				
PR	04-OCT-1999;	99US-0157117P.				
PR	05-OCT-1999;	99US-0157533P.				
PR	06-OCT-1999;	99US-0157865P.				
PR	07-OCT-1999;	99US-0158029P.				
PR	08-OCT-1999;	99US-0158232P.				
PR	12-OCT-1999;	99US-0158369P.				
PR	13-OCT-1999;	99US-0159293P.				
PR	13-OCT-1999;	99US-0159294P.				
PR	13-OCT-1999;	99US-0159295P.				
PR	14-OCT-1999;	99US-0159329P.				
PR	14-OCT-1999;	99US-0159330P.				
PR	14-OCT-1999;	99US-0159331P.				
PR	14-OCT-1999;	99US-0159637P.				
PR	14-OCT-1999;	99US-0159638P.				
PR	18-OCT-1999;	99US-0159584P.				
PR	21-OCT-1999;	99US-0160741P.				
PR	21-OCT-1999;	99US-0160767P.				
PR	21-OCT-1999;	99US-0160768P.				
PR	21-OCT-1999;	99US-0160770P.				
PR	21-OCT-1999;	99US-0160814P.				
PR	21-OCT-1999;	99US-0160815P.				
PR	22-OCT-1999;	99US-0160980P.				
PR	22-OCT-1999;	99US-0160981P.				
PR	22-OCT-1999;	99US-0160989P.				
PR	25-OCT-1999;	99US-0161404P.				
PR	25-OCT-1999;	99US-0161405P.				
PR	25-OCT-1999;	99US-0161406P.				
PR	26-OCT-1999;	99US-0161359P.				
PR	26-OCT-1999;	99US-0161360P.				
PR	26-OCT-1999;	99US-0161361P.				
PR	28-OCT-1999;	99US-0161520P.				
PR	28-OCT-1999;	99US-0161992P.				
PR	28-OCT-1999;	99US-0161993P.				
PR	28-OCT-1999;	99US-0162142P.				
Query Match						
Best Local Similarity 31.5%; Score 969.5; DB 3; Length 523;						
Matches 211; Conservative 99; Mismatches 173; Indels 49; Gaps 12;						
QY	27	IALLVVLVSLVQWLSRLKQKGRSPWPIGATVEQLNRHMDMLVGYL-----SRHRT	82			
Db	10	LTIVTVIIVFSLR-----RYKGRVWPLVGSIPALITNAHRMDFADNLRCGGTYQT	66			
QY	83	VTVDMPTS-----YTYIADPNVNHVLTNTFTNPKGIYVRSYMDVLLGDGIFNADGEL	137			
Db	67	CIFPIPLAKQGHVTVTCDPKNLEHLKTRFDNPKYKPSWQSVFHDLLGDGIFNSDGT	126			
QY	138	WRKORKTASFEFASKNLRDFA-IVFREYLSKLSGILSQASKAGKVVDQMELYNRWTLDS	196			
Db	127	WRFORKTAALFEFTTTLRQAMRWVDRANKRLVPILESARSABP:DLQVLLRLTFDN	186			
QY	197	ICKVGFGEIGTLPDLPENSFAQAFDAANIIILRLIDP--LWRIKRFHFVGVSEALLAQ	254			
Db	187	ICGLTFGRKPTLSPEFPENGFAVDFGATQATLQRFIMEFPIWKRLNLGLGEDDMSR	246			
QY	255	SIKLVDEFTYSVIRRRKAEIYEVRSAGKQEKMKH-DILSRFIELGEAGDDGGFGDDKSL	313			
Db	247	SISHVDNVLSEIINTRKLELL-----GQQQDSRHDDLRLSRPMKKES-----YSDKYL	295			

RESULT 7  
AAG24684  
ID AAG24684 standard; protein; 522 AA.  
XX  
AC AAG24684;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28454.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 14-MAY-1999; 99US-0134376P.  
PR 18-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144003P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-015753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159283P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.

PR	29-OCT-1999;	99US-0162142P.
Query Match	31.2%;	Score 961.5; DB 3; Length 522;
Best Local Similarity	39.4%;	Pred No. 3.2e-88;
Matches 209;	Conservative 98;	Mismatches 177; Indels 47; Gaps 11;
QY	27	IALLVLSWIIQVRSWRKKQKQSPWFGVATVQELRNYHRHMDLVGYL-----SRHRT 82
Db	9	LTLLVTVIIFWVSLR---RSYKGRVWPLVGLSPALITNAHRMHDFIADNLRMCGGTQYT 65
QY	83	VTVDMPTFS-----YTYIADPVNVEHLKTNFTNYPKGIIVYSYVMDVLLGCGIFNADGEL 137
Db	66	CIPPIFLAKKQGHVTVTCDPKNLEHLKTRFDNYPKGSQWSVFHDLGCGIFNDSGDT 125
QY	138	WRKORTASFEFASKNLRDFS-IVPREYSLKSLGILSOAKAGKQVMDQMELYMNTLDS 196
Db	126	WRPQRTALEFTFTRLRQAMAWVDRAIKNRLVPILESARSAPFIDLQVLLRLITFDN 185
QY	197	ICKVGFGEIGTSLPDLPENSAQAFDAANIITLRFIDP--LWRIKRFHVHVSALLAQ 254
Db	186	ICGLTFGKDPRTLSPFPENGFAVFDGATEATLQRFIMPEFIWKIRKWLRLGLEDDMSR 245
QY	255	SIKLVDEFYSVTRRKAEIIVEVRASGKQKQKHDIILSRFIELGRAGDGGGFGDDKSLR 314
Db	246	SISHVDNYLSEIINTRKLELGLQQQDQSRH---DDLLSRFMKKES-----YSDKYLK 295
QY	315	DVVLNFIAGRDTATTLSWFTHMASHPDVAEKLRRLECAFEARAREEGVTLVLCGA 374
Db	296	YVALNFIAGRTSSVAMSHFFWLVSLNPRVEKINEIC-----TILL----- 339
QY	375	DADDKAFARVAQFAG-LTYDSLGLKLVYLHACVTTTLKLYPAVPODPKGILLEDVLPDG 433
Db	340	---KTRDTNYSKMTDEPLTFDEIDQLVYLKAALSETLRLYPSVPEDSKFVVANDVLPDG 395
QY	434	TKVRAGMWTVPYSYGRMEYNHGPDAASPRPERMINEDGAPNAPSPFKETAFAQGRPIC 493
Db	396	TFVPSGNNVTYSYISVGRMKFIWGEDCLEFPPERWL-EESRDEKCMQYKFPVAFNAGRIC 454
QY	494	LGKDSAYLQMKALAILFREYSFRLLEGHPVQVYRMMTILSMAGHLKVRVSR 544
Db	455	LGKDLAYLQMKSTITASILLRHRLTVAGHVRVEQKMSLTLPFKFGLKMDVHK 505
RESULT 8		
AAG24683	ID	AAG24683 standard; protein; 527 AA.
XX	AC	AAG24683;
XX	DT	17-OCT-2000 (first entry)
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 28453.
XX	KW	Protein identification; signal transduction pathway; metabolic pathway;
XX	KW	hybridisation assay; Genetic mapping; gene expression control; promoter;
XX	KW	termination sequence.
XX	OS	Arabidopsis thaliana.
XX	FN	EP1033405-A2.
XX	PD	06-SEP-2000.
XX	PF	25-FEB-2000; 2000EP-00301439.
XX	PR	25-FEB-1999; 99US-0121825P.
PR	05-MAR-1999;	99US-0123180P.
PR	09-MAR-1999;	99US-0123548P.
PR	23-MAR-1999;	99US-0125788P.
PR	25-MAR-1999;	99US-0126264P.
PR	29-MAR-1999;	99US-0126785P.
PR	01-APR-1999;	99US-0127462P.
PR	06-APR-1999;	99US-0128234P.
PR	25-FEB-1999;	99US-0121825P.
PR	05-MAR-1999;	99US-0123180P.
PR	09-MAR-1999;	99US-0123548P.
PR	23-MAR-1999;	99US-0125788P.
PR	25-MAR-1999;	99US-0126264P.
PR	29-MAR-1999;	99US-0126785P.
PR	01-APR-1999;	99US-0127462P.
PR	06-APR-1999;	99US-0128234P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144086P.
PR	16-JUL-1999;	99US-0144086P.
PR	13-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	08-APR-1999;	99US-0128714P.
PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	28-APR-1999;	99US-0130891P.
PR	28-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	30-APR-1999;	99US-0132407P.
PR	04-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	21-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136352P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99

RESULT 9  
AAG36729  
ID AAG  
XX  
AC AAG



XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45052.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW Hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131449P.  
XX 30-APR-1999; 99US-0132048P.  
XX 04-MAY-1999; 99US-0132407P.  
XX 05-MAY-1999; 99US-0132484P.  
XX 06-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 07-MAY-1999; 99US-0132487P.  
XX 11-MAY-1999; 99US-0132863P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 04-JUN-1999; 99US-0137502P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139452P.  
XX 18-JUN-1999; 99US-0139454P.  
XX 18-JUN-1999; 99US-0139455P.  
XX 18-JUN-1999; 99US-0139456P.  
XX 18-JUN-1999; 99US-0139457P.  
XX 18-JUN-1999; 99US-0139458P.  
XX 18-JUN-1999; 99US-0139459P.  
XX 18-JUN-1999; 99US-0139460P.  
XX 18-JUN-1999; 99US-0139461P.  
XX 18-JUN-1999; 99US-0139462P.  
XX 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.

PR	27-AUG-1999;	99US-0151065P.	Db	238	KKBS-----YSDETQORVALNFILAGDRTSSVALSWFFWLTITQHPAIEDKILREICTV	290
PR	27-AUG-1999;	99US-0151066P.				
PR	27-AUG-1999;	99US-0151080P.	QY	357	EAERAREEGVTILVLCGGADADKAPAAARVAQFAGILTYDSLGKIVYLHACVTETLRLYPA	416
PR	30-AUG-1999;	99US-01511303P.				
PR	31-AUG-1999;	99US-0151438P.	Db	291	LVE-TRGDDVALW-----TDEP-----LSCELDRLVFLKAALSETLRLYPS	331
PR	01-SEP-1999;	99US-0151930P.				
PR	07-SEP-1999;	99US-01522363P.	QY	417	VPQDPKGILEDVLPDGTIKVRAGMVTVPYVSMGRMEYNWGPDAASFRRPERWINE--GA	474
PR	10-SEP-1999;	99US-0153070P.				
PR	13-SEP-1999;	99US-0153758P.	Db	332	VEEDSKRAVKDDVLPDGTFPVAGSSITVSIYAGRMKSTWGEDCIEFKPERMISQSDGGR	391
PR	15-SEP-1999;	99US-0154018P.				
PR	16-SEP-1999;	99US-0154039P.				
PR	20-SEP-1999;	99US-0154779P.	QY	475	ERNASPFKETAQAGPRICLGKXDSAYLQMK-MALAILFRFYFRLLGHPVQVRRMTIIS	533
PR	22-SEP-1999;	99US-0155139P.				
PR	23-SEP-1999;	99US-0155486P.	Db	392	FINHDPFKFVAVNAGPRICLGKDLAYLQMKSIASAVLLR-HRLTVVTGKHVEQKMSLTIF	450
PR	24-SEP-1999;	99US-0155659P.				
PR	28-SEP-1999;	99US-0156458P.	QY	534	MAHGLKRVV	542
PR	29-SEP-1999;	99US-0156596P.	Db	451	MKYGLLVNV	459
PR	04-OCT-1999;	99US-0157117P.				
PR	05-OCT-1999;	99US-0157753P.				
PR	06-OCT-1999;	99US-0157865P.				
PR	07-OCT-1999;	99US-0158029P.				
PR	08-OCT-1999;	99US-0158232P.				
PR	12-OCT-1999;	99US-0158369P.				
PR	13-OCT-1999;	99US-0159293P.				
PR	13-OCT-1999;	99US-0159294P.				
PR	13-OCT-1999;	99US-0159295P.				
PR	14-OCT-1999;	99US-0159339P.				
PR	14-OCT-1999;	99US-0159331P.				
PR	14-OCT-1999;	99US-0159637P.				
PR	14-OCT-1999;	99US-0159638P.				
PR	18-OCT-1999;	99US-0159584P.				
PR	21-OCT-1999;	99US-0160741P.				
PR	21-OCT-1999;	99US-0160767P.				
PR	21-OCT-1999;	99US-0160788P.				
PR	21-OCT-1999;	99US-0160770P.				
PR	21-OCT-1999;	99US-0160814P.				
PR	21-OCT-1999;	99US-0160815P.				
PR	22-OCT-1999;	99US-0160980P.				
PR	22-OCT-1999;	99US-0160981P.				
PR	22-OCT-1999;	99US-0160989P.				
PR	25-OCT-1999;	99US-0161404P.				
PR	25-OCT-1999;	99US-0161405P.				
PR	25-OCT-1999;	99US-0161406P.				
PR	26-OCT-1999;	99US-0161359P.				
PR	26-OCT-1999;	99US-0161360P.				
PR	26-OCT-1999;	99US-0161361P.				
PR	28-OCT-1999;	99US-0161920P.				
PR	28-OCT-1999;	99US-0161992P.				
PR	28-OCT-1999;	99US-0161993P.				
PR	29-OCT-1999;	99US-0162142P.				
Query Match 29.6%; Score 911.5; DB 3; Length 489;						
Best Local Similarity 41.9%; Pred.No.3.6e-83;						
Matches 205; Conservative 77; Mismatches 162; Indels 45; Gaps 12;						
QY	69	MHDNLVGVYL-----SRHRTVTVDMPFTS-----YTYIADPVNVHVLKTNFTNPKGIYVR	119			
Db	1	MHDNISDNLKACSGYQTCICAPFLAKKQGLVTVTCDPNLEHLKKNRFDNPKGPTWQ	60			
QY	120	SYMDVLGGIFINAGELMRKORTASPEFASKNLRDFS-AIVFREYSKLKSGILSQASK	178			
Db	61	AVFHDLGGQIFNSDGDWLFORKTAALFTTTLRQAVRWVNRKILRFLPILLENARL	120			
QY	179	AKQVDMQELNYRMILTDLSCKVGFGEICTLSPDLSPENSFAOFAAANIIITLRFIDP--	236			
Db	121	GSEPIDQLQLRLTLDNFCGLTFGDKDPRTCAPGLPVNTFAVAFDRATEASLQRIPLPI	180			
QY	237	LWKRIKFFHVGSEALLAQSIKLVDFTYSYVIRRRKAEIIVEVRASGKQERKQKHDIISRFT	296			
Db	181	LWKFKWLRGLGVSTRLSLVQDNYLSIIITRKEMMTQHNKGKH---DDLRSRFLK	237			
QY	297	LGEAGDGGGFGDKSLRDVNLVFIAGRDTTATLTSWETHMAMSHPDVAEKLRLELCAF	356			

RESULT 10

AAG54204

ID AAG54204 standard; protein; 475 AA.

XX AC AAG54204;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 69088.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EPI033405-A2.

XX FD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 05-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123588P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126264P.

XX PR 01-APR-1999; 99US-0126785P.

XX PR 06-APR-1999; 99US-0127462P.

XX PR 08-APR-1999; 99US-0128234P.

XX PR 16-APR-1999; 99US-0128714P.

XX PR 19-APR-1999; 99US-0129845P.

XX PR 21-APR-1999; 99US-0130077P.

XX PR 23-APR-1999; 99US-0130449P.

XX PR 23-APR-1999; 99US-0130510P.

XX PR 28-APR-1999; 99US-0130891P.

XX PR 30-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 04-MAY-1999; 99US-0132407P.

XX PR 05-MAY-1999; 99US-0132485P.

XX PR 06-MAY-1999; 99US-0132486P.

XX PR 06-MAY-1999; 99US-0132487P.

XX PR 07-MAY-1999; 99US-0132863P.

XX PR 11-MAY-1999; 99US-0134256P.

XX PR 14-MAY-1999; 99US-0134218P.

XX PR 14-MAY-1999; 99US-0134219P.

XX PR 14-MAY-1999; 99US-0134221P.

XX PR 14-MAY-1999; 99US-0134370P.

XX PR 18-MAY-1999; 99US-0134768P.

XX PR 19-MAY-1999; 99US-0134941P.

XX PR 20-MAY-1999; 99US-0135124P.

XX PR 21-MAY-1999; 99US-0135353P.



PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.  
Query Match 29.0%; Score 894; DB 3; Length 475;  
Best Local Similarity 40.2%; Pred. No. 2e-81;  
Matches 197; Conservative 90; Mismatches 157; Indels 46; Gaps 11;  
QY 69 MHDWLVGYL-----SRHRTVTVDMPFTS-----YTIADPVNVHVLKTNFTNYPKGIVR 119  
Db 1 MHDPIADNLRMCGTQTCTCFIPFLAKKGHVTVCDPKNLEHILKTRFDNFKGPSWQ 60  
QY 120 SYMDVLGDIYNADGELMRKQRTKTSFEPASKNLRDPSA-IVPRYSYKLSGILSQASK 178  
Db 61 SVFHDLLGDIYNFSDGDTWRFORKTALEFTTLRQAMARWYDRAIKRNLVFILESARS 120  
QY 179 AGKVDMDQELYMWTLDISCKVGFVEIGTSLPDLSPNSPAQAFDAANIITLRFIDP-- 236  
Db 121 RAEPIDLQVLLRLTFDNLICGLTFGKDPRTLSPEFFENGFAVADGATEATLQRFIMPEF 180  
QY 237 LMRKIRFFHVSGEALLAQSIKLVDEFTYSVIRRRKAEIIVEVRASGKQKMKH-DILSRFI 295  
Db 181 IWKIRKWLRLGLEDDXSRSSISHVDNYLSEIINTRKLELL-----GQQQDESRRHDDLRSFM 236  
QY 296 ELGAGDGGGFGDDKSLRDVVLNFIAGHDTTATLSWFTWAMSHPDVAEKLRELCA 355  
Db 237 KKKES-----YSDKYLYKVALNFILAGRDTSSVAMSWFFLWLSNPRVEEKIINEIC- 288  
QY 356 FEARAREEGVTVLGGADADDKAFARVAQFAG-LLTYDSLGKLVHLHACVTEILRLY 414  
Db 289 -----TILI-----KTRDTNVSKWTDEPLTDEIDQLVYKAAALSEILRLY 329  
QY 415 PAVPQPKGILEDDVLPDGTQVRAGGMVTVPYSGMRMEYNWGPDAASPERWINEEDGA 474  
Db 330 PVPEDSKFVANDVLPDGTQVPSGSMVTYSYSGRMKEIWGEDCIEFKPERWL-EESR 388  
QY 475 FRNASPKFTAFQAGPRICLGKDSAYLQMKALAILERFYSFRLLGHPVQYRMMTLSM 534  
Db 389 DEKNQYKFVAFNAGPRICLGKOLAYLQMKSIITASILLRHLTVAPGHRVEQKXSLTFM 448  
QY 535 AHGKVRVSR 544  
Db 449 KFLKMDVHK 458  
RESULT 11  
AAG24685  
ID AAG24685 standard; protein; 475 AA.  
XX  
AC AAG24685;  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28455.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 18-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140895P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.

PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144332P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144332P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144884P.	PR	18-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160989P.
PR	26-JUL-1999;	99US-0145227P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145919P.	PR	26-OCT-1999;	99US-0161359P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161360P.
PR	28-JUL-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161920P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161929P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161933P.
PR	04-AUG-1999;	99US-0147204P.	PR	29-OCT-1999;	99US-0162142P.
PR	05-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147192P.			
PR	06-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147932P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	30-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
PR	13-SEP-1999;	99US-0153758P.			
PR	15-SEP-1999;	99US-0154018P.			
PR	16-SEP-1999;	99US-0154039P.			
PR	20-SEP-1999;	99US-0154779P.			
PR	22-SEP-1999;	99US-0155139P.			
PR	23-SEP-1999;	99US-0155486P.			
PR	24-SEP-1999;	99US-0155659P.			
PR	28-SEP-1999;	99US-0156458P.			
PR	29-SEP-1999;	99US-0156596P.			
PR	04-OCT-1999;	99US-0157117P.			
PR	05-OCT-1999;	99US-0157753P.			
PR	06-OCT-1999;	99US-0157865P.			
PR	07-OCT-1999;	99US-0158029P.			
PR	08-OCT-1999;	99US-0158232P.			
PR	12-OCT-1999;	99US-0158369P.			
Query Match			28.8%; Score 886; DB 3; Length 475;		
Best Local Similarity			39.9%; Pred. No. 1.3e-80;		
Matches 195; Conservative			89; Mismatches 161; Indels 44; Gaps 10;		
QY	69	MHDWLVGL-----SRHRTVTVDMPFTS-----YTIADPNVNVHVLTNFTNPKGIVYR 119			
DB	1	MHDEIADNLRMCGGYQTCTPIPFLLAKQKHGVTVTCDPKNLEHLKTRFDNPKGSPSWQ 60			
QY	120	SYMDVLLGDGFNADGELWRKQRTAFSEFASKNLRDFA-IVREYSLKLSGLTSQASK 178			
DB	61	SVFHDLLGDGFNDSGDGTNRFORKTALEFFTRTLRQAMAEWVDRAIKRLVPILESARS 120			
QY	179	AGKVVDMOELYMRTLDSICKVGFVEIGTSLPDPENSFAQAFDAANIITLRFIDP-- 236			
DB	121	RAPIDLDQVLLRLTDFDNICGLTFGKOPRTLSPEFPENGFAVAFDGAATEALQRFIMPEF 180			
QY	237	LWRIKRPFHVGSEALLAQSIKLVDBFTYSVIRRKAEIVEVRASGQEKMKHDLISREIE 296			
DB	181	IWKIRKWLRLGLEDDMSRSISHDVNYLSEIINTRKLELLGQOQDGRH---DDLRSFMK 237			
QY	297	LGEAGDDGGFGDDKSLRDVVLNFVIAGRDATTATLSWFTHMAMSHPDVAEKLRELCAF 356			
DB	238	KKES-----YSDKYLKVALNFILAGRNTSSVAMSWFFWLVSINPRVEEKINEIC-- 288			
QY	357	EAERAREEGVTILVLCGGADADDKAPAAVAFAG-LITYDSLGLKLVYHACVTETLRLYP 415			
DB	289	-----TILI-----KTRDTNVSKWTDPLTFDEIDQIVLYLKAALSETLRLYP 330			
QY	416	AVPQDPKGILEDVLPDGTQVRAGMTVYVPSYSGRMEVNWGPDAAASPRPERWINEGCAF 475			
DB	331	SYPEDSKFVANDVLPDGTFFVPSGNSVTYSYSGVRMKFTWGEDCLEFEPKRWL-EESRD 389			
QY	476	RNASPFKFTAFQAGPRICLGKDSAYLOMKMALAIIFREYFSFLLFEGHPVQVYRMTILSMA 535			
DB	390	EKCNOYKVFVAFNAGPRICLGKDLAYLOMKSIATASILLRHRLTVAPGHRVQEQMSUTLFWK 449			
QY	536	HGLKVRVSR 544			
DB	450	FGLKMDVHK 458			
RESULT 12			AAG54205		
ID			AAG54205 standard; protein; 465 AA.		

XX	AGS54205;	18-OCT-2000	(first entry)	PR	18-JUN-1999;
XX				PR	18-JUN-1999;
XX				PR	18-JUN-1999;
XX				PR	18-JUN-1999;
DT				PR	21-JUN-1999;
DT				PR	21-JUN-1999;
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 59089.			PR	22-JUN-1999;
DE				PR	22-JUN-1999;
DE				PR	23-JUN-1999;
KW	Protein identification; signal transduction pathway; metabolic pathway;			PR	23-JUN-1999;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			PR	24-JUN-1999;
KW	termination sequence.			PR	24-JUN-1999;
KW				PR	28-JUN-1999;
XX				PR	29-JUN-1999;
OS	Arabidopsis thaliana.			PR	30-JUN-1999;
XX				PR	30-JUN-1999;
PN	EP1033405-A2.			PR	01-JUL-1999;
XX				PR	01-JUL-1999;
PD				PR	02-JUL-1999;
XX	06-SEP-2000.			PR	06-JUL-1999;
XX				PR	08-JUL-1999;
XX				PR	09-JUL-1999;
XX	25-FEB-2000; 2000EP-00301439.			PR	12-JUL-1999;
XX				PR	13-JUL-1999;
PR	25-FEB-1999;		99US-0121825P.	PR	13-JUL-1999;
PR	09-MAR-1999;		99US-0123180P.	PR	14-JUL-1999;
PR			99US-0123348P.	PR	15-JUL-1999;
PR	23-MAR-1999;		99US-0125788P.	PR	15-JUL-1999;
PR	25-MAR-1999;		99US-0126264P.	PR	16-JUL-1999;
PR	23-MAR-1999;		99US-0126785P.	PR	16-JUL-1999;
PR	01-APR-1999;		99US-0127462P.	PR	19-JUL-1999;
PR	06-APR-1999;		99US-0128234P.	PR	19-JUL-1999;
PR	08-APR-1999;		99US-0128714P.	PR	19-JUL-1999;
PR	16-APR-1999;		99US-0129845P.	PR	19-JUL-1999;
PR	19-APR-1999;		99US-0130077P.	PR	19-JUL-1999;
PR	21-APR-1999;		99US-0130049P.	PR	19-JUL-1999;
PR	23-APR-1999;		99US-0130510P.	PR	19-JUL-1999;
PR	28-APR-1999;		99US-0131449P.	PR	19-JUL-1999;
PR	30-APR-1999;		99US-0132048P.	PR	20-JUL-1999;
PR	30-APR-1999;		99US-0132240P.	PR	20-JUL-1999;
PR	04-MAY-1999;		99US-0132484P.	PR	20-JUL-1999;
PR	05-MAY-1999;		99US-0132485P.	PR	21-JUL-1999;
PR	06-MAY-1999;		99US-0132486P.	PR	21-JUL-1999;
PR	06-MAY-1999;		99US-0132487P.	PR	21-JUL-1999;
PR	07-MAY-1999;		99US-0132863P.	PR	22-JUL-1999;
PR	11-MAY-1999;		99US-0134256P.	PR	22-JUL-1999;
PR	14-MAY-1999;		99US-0134218P.	PR	22-JUL-1999;
PR	14-MAY-1999;		99US-0134219P.	PR	23-JUL-1999;
PR	14-MAY-1999;		99US-0134221P.	PR	23-JUL-1999;
PR	14-MAY-1999;		99US-0134370P.	PR	23-JUL-1999;
PR	18-MAY-1999;		99US-0134768P.	PR	26-JUL-1999;
PR	19-MAY-1999;		99US-0134941P.	PR	26-JUL-1999;
PR	20-MAY-1999;		99US-0135124P.	PR	27-JUL-1999;
PR	21-MAY-1999;		99US-0135353P.	PR	27-JUL-1999;
PR	24-MAY-1999;		99US-0135629P.	PR	27-JUL-1999;
PR	25-MAY-1999;		99US-0136021P.	PR	27-JUL-1999;
PR	27-MAY-1999;		99US-0136392P.	PR	28-JUL-1999;
PR	28-MAY-1999;		99US-0136782P.	PR	02-AUG-1999;
PR	01-JUN-1999;		99US-0137232P.	PR	02-AUG-1999;
PR	03-JUN-1999;		99US-0137588P.	PR	02-AUG-1999;
PR	04-JUN-1999;		99US-0137502P.	PR	03-AUG-1999;
PR	07-JUN-1999;		99US-0137724P.	PR	04-AUG-1999;
PR	08-JUN-1999;		99US-0138094P.	PR	04-AUG-1999;
PR	10-JUN-1999;		99US-0138540P.	PR	05-AUG-1999;
PR	10-JUN-1999;		99US-0138847P.	PR	05-AUG-1999;
PR	14-JUN-1999;		99US-0139119P.	PR	06-AUG-1999;
PR	16-JUN-1999;		99US-0139452P.	PR	06-AUG-1999;
PR	16-JUN-1999;		99US-0139453P.	PR	09-AUG-1999;
PR	17-JUN-1999;		99US-0139492P.	PR	09-AUG-1999;
PR	18-JUN-1999;		99US-013945		





us-10-021-657-2.rag

Mon Mar 8 10:13:27 2004

PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140351P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 28-SEP-1999; 99US-0156599P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.

Mon Mar 8 10:13:27 2004

us-10-021-657-2.rag

```
PR 29-OCT-1999; 99US-0162142P.
Query Match 27.5%; Score 847; DB 3; Length 508;
Best Local Similarity 36.2%; Pred. No. 1.4e-76;
Matches 195; Conservative 90; Mismatches 188; Indels 66; Gaps 12;
QY 27 IALLVVLVSWILVQVSLRKQK-----GPRSPVIGATVQQLRNYHRHMDLVGY--L 77
DB 6 ILILEPIIGVLFISPTTLAKTASPSPNPISEPLIGSILSFNKNRHLQWYTDLRL 65
QY 78 SRHRTVVDMPSTSYNI-ADPNVVEHLKTNFTNPKGIVVRSYMDVLLGDIENADGE 136
DB 66 SPQQTIVDLLFQRRTIITANPENVEHLKTNFTNPKGKPTDLLGDLGGIFNSDGE 125
QY 137 LMRKQKRTASFEFASKNLRDFAIVPR-FYSLKSLGILSOAKGKVKVDMQELYMNTLD 195
DB 126 LWSQKRLASHFTMSLREFTPEILREVEVQNLIPVLSAVDCGTVDFQVLRKFAFD 185
QY 196 STCKVGFVEIGTSLPDLFENSPAQAFDAANIITLRFIDP---LWRIKRFHVGSALL 252
DB 186 VVCKVSLGWDPCDCLDTRVPVELVKAFDVAEISARRATEPVYAVWVKRFLNVGSEKRL 245
QY 253 AOSIKLVDEFTYSVIRRRKAEIVEVRAASKOERKMDILSRFTLGEAGDDGGFGDDKS 312
DB 246 REAKTVHLSVSEIIAKKSL---DIGGVSDKQDLLSRFLAAGHG-----EEA 292
QY 313 LRDLVLFVAGDRTTATTLTSMTHMAMSHPDVAEKLRRLCAFEAREAREEGVTLVLCG 372
DB 293 VRDSVIFIMAGDRTTSAAMTFLWLLSQNDVETKILDEL-----RNKG----- 337
QY 373 GADADOKAFARVAQFAGLLTYSGLKLYLVHACVTELRLYPAVQDPKGILEDVLPD 432
DB 338 -----SLGLGFEDLRMSYTKACLCCEAMRLYPPVAVDSKHAANDLILPD 361
QY 433 GTKVRAGGVTVYPVSMGRMEYNWGPDAASFRPERWINEEDGAF-----RNASPPKFTAF 486
DB 382 GTPLKKGDKVTYPYNGRMEKVKWGDWDEFKPNRWFEEPSYGTKEPVLKSVSSKFFVF 441
QY 487 QAGPRICLGKDSXLOKMMAL-AILFRFYRLLEGHVPVYRMVMTLSMAHGLKVRVS 544
DB 442 QAGPRVICGKEMAFQMKYVGVSVLSRFKIPVCNNRPVFPVPLIT-AHMAAGGLKVKIKR 499

RESULT 14
AAG18700
ID AAG18700 standard; protein; 510 AA.
XX
AC AAG18700;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20214.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
DD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 16-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 07-MAY-1999; 99US-0134256P.
PR 11-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142377P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
```

RESULT 15  
AAG18699  
ID AAG1  
XX  
AC AAG1

XX	17-OCT-2000	(first entry)	Arabidopsis thaliana protein fragment SEQ ID NO: 20213.	PR 18-JUN-1999; 99US-0139763P; 99US-0139817P; 99US-0139861P; 99US-0139899P; 99US-0140353P; 99US-0140354P; 99US-0140655P; 99US-0140823P; 99US-0140991P; 99US-01411287P; 99US-01411842P; 99US-0142154P; 99US-0142155P; 99US-0142390P; 99US-0142803P; 99US-0142920P; 99US-0142977P; 99US-0143542P; 99US-0143624P; 99US-0144005P; 99US-0144085P; 99US-0144086P; 99US-0144325P; 99US-0144331P; 99US-0144332P; 99US-0144333P; 99US-0144334P; 99US-0144335P; 99US-0144352P; 99US-0144632P; 99US-0144884P; 99US-0144814P; 99US-0145086P; 99US-0145087P; 99US-0145089P; 99US-0145122P; 99US-0145145P; 99US-0145218P; 99US-0145224P; 99US-0145276P; 99US-0145913P; 99US-0145918P; 99US-0145919P; 99US-0145951P; 99US-0146386P; 99US-0146388P; 99US-0147038P; 99US-0147204P; 99US-0147302P; 99US-0147192P; 99US-0147260P; 99US-0147303P; 99US-0147416P; 99US-0147493P; 99US-0147935P; 99US-0148171P; 99US-0148319P; 99US-0148341P; 99US-0148565P; 99US-0148684P; 99US-0149175P; 99US-0149368P; 99US-0149426P; 99US-0149722P; 99US-0149723P; 99US-0149929P; 99US-0149902P; 99US-0149930P; 99US-0150566P; 99US-0150884P; 99US-0150885P; 99US-0150886P; 99US-0150887P; 99US-0150888P; 99US-0150889P; 99US-0150890P; 99US-0150891P; 99US-0150892P; 99US-0150893P; 99US-0150894P; 99US-0150895P; 99US-0150896P; 99US-0150897P; 99US-0150898P; 99US-0150899P; 99US-0150900P; 99US-0150901P; 99US-0150902P; 99US-0150903P; 99US-0150904P; 99US-0150905P; 99US-0150906P; 99US-0150907P; 99US-0150908P; 99US-0150909P; 99US-0150910P; 99US-0150911P; 99US-0150912P; 99US-0150913P; 99US-0150914P; 99US-0150915P; 99US-0150916P; 99US-0150917P; 99US-0150918P; 99US-0150919P; 99US-0150920P; 99US-0150921P; 99US-0150922P; 99US-0150923P; 99US-0150924P; 99US-0150925P; 99US-0150926P; 99US-0150927P; 99US-0150928P; 99US-0150929P; 99US-0150930P; 99US-0150931P; 99US-0150932P; 99US-0150933P; 99US-0150934P; 99US-0150935P; 99US-0150936P; 99US-0150937P; 99US-0150938P; 99US-0150939P; 99US-0150940P; 99US-0150941P; 99US-0150942P; 99US-0150943P; 99US-0150944P; 99US-0150945P; 99US-0150946P; 99US-0150947P; 99US-0150948P; 99US-0150949P; 99US-0150950P; 99US-0150951P; 99US-0150952P; 99US-0150953P; 99US-0150954P; 99US-0150955P; 99US-0150956P; 99US-0150957P; 99US-0150958P; 99US-0150959P; 99US-0150960P; 99US-0150961P; 99US-0150962P; 99US-0150963P; 99US-0150964P; 99US-0150965P; 99US-0150966P; 99US-0150967P; 99US-0150968P; 99US-0150969P; 99US-0150970P; 99US-0150971P; 99US-0150972P; 99US-0150973P; 99US-0150974P; 99US-0150975P; 99US-0150976P; 99US-0150977P; 99US-0150978P; 99US-0150979P; 99US-0150980P; 99US-0150981P; 99US-0150982P; 99US-0150983P; 99US-0150984P; 99US-0150985P; 99US-0150986P; 99US-0150987P; 99US-0150988P; 99US-0150989P; 99US-0150990P; 99US-0150991P; 99US-0150992P; 99US-0150993P; 99US-0150994P; 99US-0150995P; 99US-0150996P; 99US-0150997P; 99US-0150998P; 99US-0150999P; 99US-0151000P; 99US-0151001P; 99US-0151002P; 99US-0151003P; 99US-0151004P; 99US-0151005P; 99US-0151006P; 99US-0151007P; 99US-0151008P; 99US-0151009P; 99US-0151010P; 99US-0151011P; 99US-0151012P; 99US-0151013P; 99US-0151014P; 99US-0151015P; 99US-0151016P; 99US-0151017P; 99US-0151018P; 99US-0151019P; 99US-0151020P; 99US-0151021P; 99US-0151022P; 99US-0151023P; 99US-0151024P; 99US-0151025P; 99US-0151026P; 99US-0151027P; 99US-0151028P; 99US-0151029P; 99US-0151030P; 99US-0151031P; 99US-0151032P; 99US-0151033P; 99US-0151034P; 99US-0151035P; 99US-0151036P; 99US-0151037P; 99US-0151038P; 99US-0151039P; 99US-0151040P; 99US-0151041P; 99US-0151042P; 99US-0151043P; 99US-0151044P; 99US-0151045P; 99US-0151046P; 99US-0151047P; 99US-0151048P; 99US-0151049P; 99US-0151050P; 99US-0151051P; 99US-0151052P; 99US-0151053P; 99US-0151054P; 99US-0151055P; 99US-0151056P; 99US-0151057P; 99US-0151058P; 99US-0151059P; 99US-0151060P; 99US-0151061P; 99US-0151062P; 99US-0151063P; 99US-0151064P; 99US-0151065P; 99US-0151066P; 99US-0151067P; 99US-0151068P; 99US-0151069P; 99US-0151070P; 99US-0151071P; 99US-0151072P; 99US-0151073P; 99US-0151074P; 99US-0151075P; 99US-0151076P; 99US-0151077P; 99US-0151078P; 99US-0151079P; 99US-0151080P; 99US-0151081P; 99US-0151082P; 99US-0151083P; 99US-0151084P; 99US-0151085P; 99US-0151086P; 99US-0151087P; 99US-0151088P; 99US-0151089P; 99US-0151090P; 99US-0151091P; 99US-0151092P; 99US-0151093P; 99US-0151094P; 99US-0151095P; 99US-0151096P; 99US-0151097P; 99US-0151098P; 99US-0151099P; 99US-0151100P; 99US-0151101P; 99US-015
----	-------------	---------------	---	---

PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 27.5%; Score 847; DB 3; Length 528;  
Best Local Similarity 26.2%; Pred. No. 1.5e-76;  
Matches 195; Conservative 90; Mismatches 188; Indels 66; Gaps 12;

Qy 27 IALLVVLVILVQWLSLRKQ-----GPRSNPVGATVQOLRNVHRMHDNLVGY--L 77  
Db 26 ILILFPIIGVILFSPFTKTKAKTASPSNFIISYPLIGSILSFNKNRHLQWYTDLLRL 85

Qy 78 SRRTVTVDMPFTSYI-ADPUNVEHVLKTNFTNPKGIVYRSYMDVLGDGIFNADGE 136  
Db 86 SPSQTIIVDLFGRRITITANPENVEHILKTNFTNFKGKPFPTDLLGDLGGGIFNSDGE 145

Qy 137 LWRKQRTASPEKALRDPFSAIVR-EYSLKLSGILSOAKAGKVVDMQELVMRMTLD 195  
Db 146 LWSQRKLASHEFTMRSLREFTFELREEVQNRLLIPLVLSAVDCGEVDFQVLKRAFD 205

Qy 196 SICKVGFGEIGTILSPDLPENSFAQAFDAANIITLRFIDP---LWRIKFFHVHVSALL 252  
Db 206 VVCKVSLGWDPCDLTLTRFVPELVKAFDVAEISARRATEPVYAVKVKRFLNVGSEKRL 265

Qy 253 AQSIKLVDEFTYSVIRRKAEIIVRASGKOEKMKHILSRFIELGAGDDGGGFGDDKS 312

Db 266 REAKTVHLSVSEIRAKKSL-----DIGGDVSDKQDLSRFLAAGHG-----EEA 312  
Qy 313 LRDVVLNFIAGRDTTATTLNFWTHMAMSHPDVAEKLERELCAFEAEARBEGVTLVLCG 372  
Db 313 VRDSVISFIMAGRDTTSAAMTWLFWLLSQNDVETKILDEL-----RNKG----- 357  
Qy 373 GADADDKAFARVAQFAGLLTYDSILGLVYLHACVTETRLYPVPQDPKGILEDVLPD 432  
Db 358 -----SIGLGFEDLREMSYTKACICEAMRLYPVAVWDSKHAANDDILPD 401  
Qy 433 GTKVRAAGMVTVPYSMGRMEYNWGPDAASFRPERWINEGAF-----RNASPKFTAF 486  
Db 402 GTPLEKGGKVTYFFPYGMEKMEKWKWDDEKPNRPFEEPSYGTKPVLSVSSFKFPVF 461  
Qy 487 QAGPRICLGKDSAYLQMKAL-AILFRFYSRLLEGHVPQYRMMTILSMAGHLKVRVR 544  
Db 462 QAGPRVCIGKEMAFQMKYVGVSVLSRFLKIPVCNNRPVFPVPLLT-AHMAGGLKVKIKR 519

Search completed: March 2, 2004, 15:27:43  
Job time : 63 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2004, 15:25:49 ; Search time 21 Seconds  
(without alignments)  
2693.363 Million cell updates/sec

Title: US-10-021-657-2

Perfect score: 3079  
Sequence: 1 EFGTREMLTPATPSFFPL.....ICVHACKGRWVSLVAVLKP 588

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	994.5	32.3	553	T01535	probable cytochrom
2	977.5	31.7	511	T00864	cytochrome P450 ho
3	971.5	31.6	537	T02450	probable cytochrom
4	970	31.5	554	G86146	hypothetical prote
5	969.5	31.5	523	B96662	probable cytochrom
6	907.5	29.5	490	T00404	cytochrome P450-11
7	904.5	29.4	550	T50510	cytochrome P450-11
8	879.5	28.6	499	T49064	cytochrome P450 CY
9	876.5	28.5	513	T85965	protein F2119.20
10	837	27.2	522	B86379	cytochrome P450-11
11	823	26.7	506	T46196	cytochrome P450-11
12	788	25.6	529	G86265	F3F19.17 protein -
13	785.5	25.5	514	T08014	cytochrome P450 CY
14	780	25.3	519	G86265	hypothetical prote
15	773	25.1	495	G84675	probable cytochrom
16	744.5	24.2	516	T09364	cytochrome P450 ho
17	744	24.2	516	T00514	cytochrome P450 ho
18	738.5	24.0	510	F84606	probable cytochrom
19	730	23.7	479	T09365	cytochrome P450 ho
20	727.5	23.6	506	T04628	probable cytochrom
21	718.5	23.3	497	G96611	cytochrome P450 ho
22	699	22.7	508	T09367	cytochrome P450 ho
23	682	22.2	503	T02357	cytochrome P450 ho
24	675	21.9	512	G96517	hypothetical prote
25	671	21.8	520	B96517	hypothetical prote
26	650	21.1	469	T09366	cytochrome P450 ho
27	619.5	20.1	480	T48311	cytochrome P450 52
28	427	13.9	519	S69988	unspecific monooxy
29	411	13.3	519	S69989	unspecific monooxy

30	408	13.3	593	2	F86441	probable cytochrom
31	407.5	13.2	523	1	O4CKA3	cytochrome P450 52
32	402	13.1	524	1	S22972	cytochrome P450 52
33	400.5	13.0	523	1	J01039	cytochrome P450 52
34	399	13.0	526	1	A40576	cytochrome P450 AL
35	392	12.7	538	1	B40576	cytochrome P450 AL
36	387	12.6	519	1	J50726	cytochrome P450 AL
37	386.5	12.6	524	2	JC7594	cytochrome P450 en
38	386.5	12.6	524	2	JC7598	cytochrome P450 en
39	385	12.5	538	1	O4CKA4	cytochrome P450 52
40	384.5	12.5	519	1	J50725	cytochrome P450 AL
41	383.5	12.5	520	2	T24780	hypothetical prote
42	381	12.4	522	1	JT0980	cytochrome P450 52
43	380.5	12.4	518	2	T24783	hypothetical prote
44	374	12.1	544	2	T27750	hypothetical prote
45	371	12.0	520	1	A46661	leukotriene B4 ome

ALIGNMENTS

RESULT 1

T01535  
probable cytochrome P450 A IG005110.21 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cross)  
C;Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 28-Jul-2000  
C;Accession: T01535  
R;Andrews, S.  
submitted to the EMBL Data Library, July 1997  
A;Description: The sequence of A. thaliana IG005110.  
A;Reference number: Z14347  
A;Accession: T01535  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-553 <AND>  
A;Cross-references: EMBL:AF013293; NID:g2252823; PID:g2252844  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Map position: 4  
A;Introns: 141/2  
A;Note: A IG005110.21  
C;Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
F;304-481/Domain: cytochrome P450 homology <P45>

Query Match	32.3%	Score	994.5	DB	2	Length	553
Best Local Similarity	41.0%	Pred. No.	3.9e-69				
Matches	226	Conservative	92	Mismatches	182	Indels	51
Gaps	15						
QY	29	LLLVLS--WILVQWLSLRKQGRSPWPVIGATVEQLNRYHRMHDWLVGYL----	SRHRT	82			
Db	8	LLVAVVAAYWLFQRLS--RWLKGPRVWVPLGSLGLEQDRMHDWITENLRACGGTYQT	66				
QY	83	VTVDMPFTS-----YTIADPVNVVHVLTNFTNYPKGIYVRSYMDVLLGGGIFNADGEL	137				
Db	67	CICAVFPLACKQGLVTVTCDEPKNIEHMLKTRFDNYPKGTQWAFVDFLQGGIFNSDGT	126				
QY	138	WRQKRTASPEFASKNLRD--FSAIVFREYSLSLGSILSQASKAGKYVDMQELMYRWLTDS	196				
Db	127	WLFQRTAALEFTRTLRQANGRWARGIKLRFCEPILFQNNVPEVVDQLILRLTFDN	186				
QY	197	ICKVGFGEVGTLSPLDPENSAQAFDAANIITLRFIDP--LWRIKRFHVSGSEALLAQ	254				
Db	187	ICGLAFKQDTRTCAPGLPENGASAFDRATEASLQRFILFEFLWRLKKWLGLEVLSLR	246				
QY	255	SIKLVDEFTYSVIRRRKAEIVFVRAGQKQMKH--DILSRFIELGEAGDGGGGGDDKSL	313				
Db	247	SLGIDGYLDVAVINTRKQLLSQRESGVQ---RHDDLRSRFMKKKDQS-----YSETFL	297				
QY	314	RDVVLNVFIAGRDTATLTSWFTHMASHDDVAEKLRELCAPFAERAREEGVTVLVCGG	373				
Db	298	RHVALNFILAGRTSSVALSFFWFLITHTPTVEDKIVREICS-----VLIETR	346				
QY	374	ADADKAFARVAQFAGLITVDSLGKLVHLHACVTETLRLYPVAPVQDPKGILEDVLPDG	433				

```

Db 347 TDVS-----SWTAEPLEFDEVDRLVYLKAALSETLRLVPSVPDSKHWVNDLIPDG 398
Qy 434 TKVAGMWTVYVYSMGRMEYNWGPDAASPRPERWIN-EDGAFNAPSPFFKETAQAGPRI 492
Db 399 TFVPAGSSVTYSIYAAGRMKSTWGEDCLFEPKPERWISPDGDKFVNHQDYRFVAFNAGPRI 458
Qy 493 CLGKDSAYLOWK-MALATLFRFYFRLLGHPVQYRMWMTILSMAHGLKVRVSRAVCHGDL 551
Db 459 CLGKDLAYLOWKTTAAVLLR-HRLTVAPGHKVECKMSLTLFMKNGLLVNVHKE----DL 513
Qy 552 DMDIVPLNPRQ 562
Db 514 EVMKSLVPKE 524

RESULT 2
T00864
cytochrome P450 homolog F17K2.4 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00864; D8491
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84891
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <STO>
A:Cross-references: GB:AE002093; NID:G2979544; PIDN:AAC06153.1; GSPDB:GNO00139
C:Genetics:
A:Gene: At2G45510; F17K2.4
A:Map position: 2
A:Introns: 174/3; 269/3; 382/3; 449/3
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology
F:302-477/Domain: cytochrome P450 homology <P45>

Query Match 31.7%; Score 977.5; DB 2; Length 511;
Best Local Similarity 41.7%; Pred. No. 7.2e-68;
Matches 215; Conservative 85; Mismatches 183; Indels 33; Gaps 7;

Qy 27 IALLVLSWILVQWSLRKQKPRSPVIGATVEQIRNHYRHMHWLGYLSRHRIVTD 86
Db 15 IFTVLCTIYIRIFETIRSRNDKRYAPVHATVFDLLFHSDELYDETEIAEKPTYEL 74
Qy 87 MPFTSYTIADPVNVFHLKTNFTNPKGIVYRSYMDVLLGDGIFNADGELWRKQRTAS 146
Db 75 SPQSEILTADPNVBEHLKTRFDNYSKGHSSENWADLLGHGIFAVDGEKWRQKLS 134
Qy 147 FEFASKNLRDFSIAIVREYSLKSLGSLQSASKAGKVVMDQELYMRTLSICKVGFVEI 206
Db 135 FEFSTRVLRDFSCSVFRNASKLVGVSEFALSKAFADQDLMLRCRTLDSIPKVGVEL 194
Qy 207 GTLSPLPE-NSFAQAFDAANIITLRFIDPLWRKIRKFFHVSGEALLAQSIKLVDEFTYS 265
Db 195 KCLDGFSGEGEFMEAFDEGNVATSRFIDPLWKLWFFNIGSQSKLKSIATIDKFVS 254
Qy 266 VTRRRKAEIVVEASGQKQKMKHDIILSRFIELGAGDGGGFGDKSLRDVILNFVIAGR 325
Db 255 LITTKREL-----AKEQNTVVRDILSRFELVESEKDPEN---MNDKYLRLDIINFVIAGK 307

```

```

Qy 326 DTTATTLSTWFTNHAMSHDPVAEKLREL--CAFEARAREEGVTLLVCGADADDKAFAA 393
Db 308 DTTAALLSFLVLMCKNPLVQEKIVQIRDVTFESHEKTTD----- 347
Qy 384 RVAQFAGLLTYDSLGLKLVHLACVTTETLRLYPAVPODPKGILLEDVLPDQTKVRAGGMVT 443
Db 348 -VNGFVESINEEALDEMHYLAALSETLRLYPPVDMRCAENDVLPDGHVRVSKGDIY 406
Qy 444 YVYSMGRMEYNWGPDAASPRPERWINEDEGAFNAPSPFFKETAQAGPRICLGKDSAYLOW 503
Db 407 YIAYAMGRWTVYVYSMGRMEYNWGPDAASPRPERWINEDEGAFNAPSPFFKETAQAGPRICLGKDSAYLOW 465
Qy 504 XMALAILFRFYSRLL-EGHPVQYRMWMTILSMAHGL 538
Db 466 KIVSMALLHFFRFKMADENSKVYKRMMLTLHVDGGL 501

RESULT 3
T02450
probable cytochrome P450 F4118.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02450; B84897
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC F4118 genomic sequence.
A:Reference number: Z14674
A:Accession: T02450
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-537 <ROU>
A:Cross-references: EMBL:AC004665; NID:G3386593; PID:G3386598
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-537 <STO>
A:Cross-references: GB:AE002093; NID:G6598466; PIDN:AAC62873.2; GSPDB:GNO0139
C:Genetics:
A:Gene: At2G45970; F4118.5
A:Map position: 2
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology
F:302-480/Domain: cytochrome P450 homology <P45>

Query Match 31.6%; Score 971.5; DB 2; Length 537;
Best Local Similarity 40.9%; Pred. No. 2.3e-67;
Matches 217; Conservative 87; Mismatches 182; Indels 45; Gaps 12;

Qy 27 IALLVLSWILVQWSLRKQKPRSPVIGATVEQIRNHYRHMHWLGYLSRHRIVTD 82
Db 7 IMLSAITAYFLWLTFISRCCLKPRVWILSLPGLIENCERHMDWISDNLACSGTYQT 66
Qy 83 VTVDMPFTS-----YTIADPVNVFHLKTNFTNPKGIVYRSYMDVLLGDGIFNADGEL 137
Db 67 CICALPFLAKKQGLVTVCDPNLEHLKRNFDNYPKGTQWQAVPHDLLQGIFNSDGT 126
Qy 138 WKQKRTASFEFASKNLRDFSIAIVREYSLKSLGSLQSASKAGKVVMDQELYMRTLS 196
Db 127 WLFQKTAALFETTRTLRQAMARWVNRALKRLFLFLENARLGSEPIQLDILLTLDN 186
Qy 197 ICKVGFVEIGTSLSPDENSEFAQAFDAANIITLRFIDP--LWIRKFFHVSGEALLAQ 254
Db 187 ICGLTFGDKPRTCPGLPVNTFAVADPDATEASLQRFILPEILWFKRWLRGLVSLTR 246
Qy 255 SIKLVDFTYSVTRRRKAEIVVEASGQKQKMKHDIILSRFIELGAGDGGGFGDKSLR 314

```



Db 247 SLVQVDNYLSEIITTKKEEMTQNNKHH---DDLSPRIKKES-----YSDETIQ 296  
 QY 315 DVVNLNFIAGROTTATTLTGWFTTHAMSHDPVAEKLRLRELCAPAEARAREEGVTLVLCGA 374  
 Db 297 RVALNFIILAGRTSSVALSWFFLIQHPAIEDKILREICTVLVE-TRGDDVALM----- 350  
 QY 375 DADKAFARVAQFAGLLTYDSLGLKLYLHACVTEITRLYPVAPQPKGILEDDVLPDGT 434  
 Db 351 -TDEP-----LSCEELDRLVFLKAALSETRLYPSPVPDSKRAVKDDVLPDGT 397  
 QY 435 KVRAGGVTVPYSGRMENWGPDAASPRPERWINE--GAFRNASPFKFTAFQAGPRI 492  
 Db 398 FVPAGSSITYISAGRMKSTWGEDCLEFPERWISQSGGRFINHDPFKFVAFNAGPRI 457  
 QY 493 CLGKDSAYLQMK-MALAILFRFYFRLLGHPVQYRMNTILSMAHGLKVRV 542  
 Db 458 CLGKDLAYLQMKSIASAVLLR-HRLTVVTGKHVEQKMSLTFMKYGLLVNV 507  
 RESULT 4  
 G86146  
 hypothetical protein F2214.14 [imported] - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C;Accession: G86146  
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Hughes, B.; Huizar, L.  
 A;Authors: Salzb, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A;Reference number: A86141; MUID:21016719; PMID:11130712  
 A;Accession: G86146  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-554 <STO>  
 A;Cross-references: GB:AF005172; NID:g920596; PIDN:AAF81318.1; GSPDB:GN00141  
 C;Genetics:  
 A;Map position: 1  
 C;Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

Query Match 31.5%; Score 970; DB 2; Length 554;  
 Best Local Similarity 40.8%; Pred. No. 3.1e-67;  
 Matches 220; Conservative 86; Mismatches 169; Indels 44; Gaps 13;

QY 28 ALLLVVL---SWILVQWLSLRKQKGRSPWVIGATVQQLRNHVRHMDLVGLY-----SRH 80  
 Db 6 AMLLVAVITGYLWLFKGIS-RWLKGRVWPLVGLSLPLGLEIQRDRHMEITENRACGGTY 64  
 QY 81 RTVVDMPFTS-----YTYIADPVNVVHLKTNFTNPKGIVYRSYMDVLLGDGIFNADG 135  
 Db 65 QTCIFAPVFLAKKGGLVITVCDPKNLEHLKTRFDNPKGSPWQSVFHDLLGQIFNSDG 124  
 QY 136 ELWRKQRTASFEFASKNLRD-FSAIVEREYSLKLSGILSQASKAGKVVDVMOELYMRTL 194  
 Db 125 DTWLFQRTAALEFTRTLRQAMGRVNRGKILRFPCLPILATAQNAEPVDLQDLILRLTF 184  
 QY 195 DSIKVGVEIGTSLPDLSPENSFAQADAAIIITLRFIDP--LWRIKRFHHVGSALL 252  
 Db 185 DNICLAFGKQTRTCAPGLPENGFAFADRAEASLQRFIIPKFMWKLKWLGLGLEVSL 244  
 QY 253 AQSILKVDFTYSVIRRRKAEIVVEVRAASKQKMKHDLISRFIELGEAGDDGGFGDDKS 312  
 Db 245 SRSGEIDYLAUVNTRKQLMSQESGTHOR-HDDLSPFMKKTES-----YSTTF 297  
 QY 313 LRDVVLNFIAGROTTATTLTGWFTTHAMSHDPVAEKLRLRELCAPAEARAREEGVTLVLCG 372  
 Db 298 LQHVALLFIILAGRTSSVALSWFFLIQHPAIEDKILREICTVLVE-TRGDDVALM-----VLIE 344

QY 373 GADADDKAFARVAQFAGLLTYDSLGLKLYLHACVTEITRLYPVAPQPKGILEDDVLPD 432  
 Db 345 TRGTDDVASWTSEP-----LGFEELDRLVFLKAALSETRLYPSPVPDSKRVENDVLPD 399  
 QY 433 GTRKVRAGGVTVPYSGRMENWGPDAASPRPERWINE-DGAFRNASPFKFTAFQAGPRI 491  
 Db 400 GTFVPAAGSVTYSIYAAGRMKSTWGEDCLEFPERWISPIDGKFINHDOYRFVAFNAGPRI 459  
 QY 492 ICLGKDSAYLQMK-MALAILFRFYFRLLGHPVQYRMNTILSMAHGLKVRVSRVCHG 549  
 Db 460 ICLGKDLAYLQMKTTAAAVLLR-HRLTVVPGHKVEQKMSLTFMKNGLLVNLKYRDLQ 517  
 RESULT 5  
 B96662  
 probable cytochrome P450 F24D7.10 [imported] - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C;Accession: B96662  
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C  
 Nature 408, 816-820, 2000  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A;Authors: Salzb, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A;Reference number: A86141; MUID:21016719; PMID:11130712  
 A;Accession: B96662  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-523 <STO>  
 A;Cross-references: GB:AF005173; NID:g6456162; PIDN:AAF09150.1; GSPDB:GN00141  
 C;Genetics:  
 A;Gene: F24D7.10  
 A;Map position: 1  
 C;Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

Query Match 31.5%; Score 969.5; DB 2; Length 523;  
 Best Local Similarity 39.7%; Pred. No. 3.1e-67;  
 Matches 211; Conservative 99; Mismatches 173; Indels 49; Gaps 12;

QY 27 IALLLVVLWILVQWLSLRKQKGRSPWVIGATVQQLRNHVRHMDLVGLY-----SHRT 82  
 Db 10 LTLTVTVIIFVSLR--RSYKGRVWPLVGLSLPALITNAHRMDFIADNLRMCQGYQT 66  
 QY 83 VTVDMPFTS-----YTYIADPVNVVHLKTNFTNPKGIVYRSYMDVLLGDGIFNADGEL 137  
 Db 67 CTFPIPLAKKGQHVITVCDPKNLEHLKTRFDNPKGSPWQSVFHDLLGQIFNSDGT 126  
 QY 138 WRKQRTASFEFASKNLRD-FSA-IVFREYSLKLSGILSQASKAGKVVDVMOELYMRTLDS 196  
 Db 127 WRFQRTAALEFTRTLRQAMGRVNRGKILRFPCLPILATAQNAEPVDLQDLVLRFLFDN 186  
 QY 197 ICKVGVEIGTSLPDLSPENSEAQADAAIIITLRFIDP--LWRIKRFHHVGSALLA 254  
 Db 187 ICGTLFGKQRTSLPFPENGFAVAFQATQATQRIIMFEFIWKIRKWLRLGLEDDMSR 246  
 QY 255 SIKLVDEFTYSVIRRRKAEIVVEVRAASKQKMKH-DILSRFIELGEAGDDGGFGDDKSL 313  
 Db 247 SIHVNDYLSFETINRKLLEL---GQQDESRRHDDLISRFMKKES-----YSKYL 295  
 QY 314 RDVVLNFIAGROTTATTLTGWFTTHAMSHDPVAEKLRLRELCAPAEARAREEGVTLVLCG 373  
 Db 296 KYVALNFIILAGRTSSVALSWFFLIQHPAIEDKILREICTVLVE-TRGDDVALM-----TILI--- 340  
 QY 374 ADADKAFARVAQFAG-LLTYDSLGLKLYLHACVTEITRLYPVAPQPKGILEDDVLPD 432  
 Db 341 -----KTRDNTVSKVDEPLTFDEIDQLVFLKAALSETRLYPSPVPDSKRVENDVLPD 395  
 QY 433 GTRKVRAGGVTVPYSGRMENWGPDAASPRPERWINE-DGAFRNASPFKFTAFQAGPRI 492

Db 396 GTFVPSGNTVTSYSGRMKFIWGEDCLEPKPERNL-EESRDEKCNQYKFVAFNAGPRI 454  
Qy 493 CLGKDSAYLQMKALAILFRYSFRLLGHPVQVRMTILSMAGHLKVRVSR 544  
Db 455 CLGKDLAYLQMKSTASILLRLHRTVAPGHRVQRQSLTLFMKGLKMDVHK 506

RESULT 6  
T00404  
Probable cytochrome P450 At2g44890 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 02-Feb-2001  
C:Accession: T00404; A84884  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Masc  
submitted to the EMBL Data Library, July 1997  
A:Description: Arabidopsis thaliana chromosome II BAC Tl3E15 genomic sequence.  
A:Reference number: Z14146  
A:Accession: T00404  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-490 <ROU>  
A:Cross-references: EMBL:AC02388; NID:G3420042; PID:G2344895  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.V.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-769, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUIB:20083487; PMID:10617197  
A:Accession: A84884  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-490 <STO>  
A:Cross-references: GB:AE002093; NID:G2344895; PID:AA031835.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g44890; Tl3E15.10  
A:Map position: 2  
A:Introns: 33/2; 153/3; 361/3; 428/3  
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
F:281-456/Domain: cytochrome P450 homology <P45>

Query Match 29.5%; Score 907.5; DB 2; Length 490;  
Best Local Similarity 40.7%; Pred. No. 1.8e-62;  
Matches 193; Conservative 85; Mismatches 167; Indels 29; Gaps 6;  
Qy 67 HEMHDLVGLSRHRTVTVDMPFTSYTADPNVHVLKTNFTNPKGIVRSYMDVLL 126  
Db 34 HKLYDTEIARTKTFRFLSPGQSEIFTADPNVHVLKTRFHNYSKPGVTNADLL 93  
Qy 127 GDGIFNADGELWRKQRTASPFASKNLRDFAIVFREYSLSKLSILSQASKAGKVDVQ 186  
Db 94 GHGIFAVDGEKWKQKLVSPFESTRVLNFSYVFTSASKLVGFIAEFALSGKSFDFQ 153  
Qy 187 ELYMRTLDLSICKGFGVIGTISPDLP-NSPAQADAAIIITLRTIDPLWR.KRFH 245  
Db 154 DMLMKCTLDLIFKVGFGVGLGDLGFSKEGEBFMKAFDEGNAGTSRVTDPFWKLKCFIN 213  
Qy 246 VGEALLAQSIKLIDEFTYSVIRRRKAEIVFVRSAGKQKMKHDLSLFIELGEAGDDGG 305  
Db 214 IGSERLKKSIALIDKFFVSLITTRKEL----SKEQNTSVREDLSKFLLESEKDPEN- 268  
Qy 306 GFDDKSLRDVNLNFVIAGRDATTATLSWFTTHAMSHPDVAEKRLRELCAPFAERAREEG 365  
Db 269 --MNDKYLRIILNVVAGKDTTAASLSWFLYMLCKNPLVQEKIVQETRDVTSSEKTD 326  
Qy 366 VTILVCGGADADDKAFARVAQFAGLLTYDSLGKLVYLHACVETILRYPAVPQPKGIL 425  
Db 327 VN-----GFIESVTEEAQMQYLHAALSETMRLYPPYPMRCAE 367  
Qy 426 EDDVLDPGTVRAGGMVTVPVSMGMMEYNWGPDAASPRPERWINEGDAFRNASPFKFTA 485  
Db 368 NDDVLDPGHRVSKGDNIIYISVANGRWITWINGQDAEEFKPERWL-KDGVFPESQFKFIS 426

Qy 486 FORGPRICIGKDSAYLQMKALAILFRYSFRLL-EGHPVOYVRMTILSMAGHL 538  
Db 427 FHAGPRICIGKDFAYRQMKIVSMALLHFFFRKMADENSKVSYKKMLTLTHVDGGL 480  
RESULT 7  
T50510  
cytochrome P450-like protein - Arabidopsis thaliana  
N:Alternate names: protein T22D6.190  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 16-Feb-2001  
C:Accession: T50510  
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De  
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25101  
A:Accession: T50510  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-550 <BEV>  
A:Cross-references: EMBL:AL357612  
A:Experimental source: cultivar Columbia; BAC clone T22D6  
C:Genetics:  
A:Map position: 5  
A:Introns: 417/3  
A:Note: T22D6.190  
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
F:330-508/Domain: cytochrome P450 homology <P45>

Query Match 29.4%; Score 904.5; DB 2; Length 550;  
Best Local Similarity 39.4%; Pred. No. 3.7e-62;  
Matches 212; Conservative 97; Mismatches 182; Indels 47; Gaps 17;  
Qy 29 LLLVVLGSLVQLVQWLSLRKQK--GPRSWFVIQ---ATVEQLRNHYRHMHWLVGL-SRRHT 82  
Db 40 LELFIAFVFATIHSLRQKKHQGVPWFLVGLMFLSLISAVRS--NIVELSDVLISQNGT 97  
Qy 83 VTVDMPFTS---YTYIADPNVHVLKTNFTNPKGIVRSYMDVLLGDGIFNADGELWR 139  
Db 98 FRFRGPFSTLNCVTCDPNVEHLKTRFSIYKGSYFRETMDLLGDLGIFNTDDGTWQ 157  
Qy 140 KQRTASPEFASKNLRDFAIVFREYSLSKLSILSQASKAGKVDVMDLYMRTLDLSICK 199  
Db 158 RQKAAASVEFHSAPRQITQSLSLHE--LVHNLRLPVLETSGK-IDLQILRLTFDNCVM 214  
Qy 200 VGFGEIGTSLSPDLSPENSFAQAFDAANIITLRTIDP--LWRIKRFPHVGSALLAOSIK 257  
Db 215 IAFGVDPGLSPKLPPIPAKAFADATEATVVRVMPKFWVKLMRSNLNLTGTEKKLESIN 274  
Qy 258 LVDEFTYSVIRRRKAEIVFVRSAGKQKMKH-DILSRFIELGEAGDGGGFGDDKSLRDV 316  
Db 275 GVDDFAFEBVIRTKEM-----SLETEIAPRDLTLTFMGL---RDENGQKFSKFLRDI 326  
Qy 317 VLVFIAGRDYTTATLSWFTTHAMSHPDVAEKRLRELCAPFAERAREEGVTILVCGGADA 376  
Db 327 CVNFIAGRDYTTATLSWFTTHAMSHPDVAEKRLRELCAPFAERAREEGVTILVCGGADA 371  
Qy 377 DDKAFARVAQFAGLLTYDSLGKLVYLHACVETILRYPAVPQPKGILEDDVLDPGTVK 436  
Db 372 -DHGDTKKNMEYEPVFERPEEIKKMDYLQAALSETLRLYPSVVDHKEVLEDDVDPDGTKL 430  
Qy 437 RAGGMVTVPVSMGMMEYNWGPDAASPRPERWINEGDAFRNASPFKFTAFQAGRICLIG 496  
Db 431 KKGSKVIYAIYANGRMETWINGKDCREKPERWL-RDGRYMSSEYAKFTAFNGGPRCLIGK 489  
Qy 497 DSAYLQVK-MALAILFRYSFRILE--GHPVOYVRMTILSMAGHLKVR-VSRVCHGD 550  
Db 490 DFAYQMRVAAAIYR-YKVRVDKGGKVKPEPKMALTNYMKHGLKVNVRVRSSEID 546

RESULT 8  
T49064

Cytochrome P450-like protein - Arabidopsis thaliana  
 N:Alternate names: protein TSP19.280  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 16-Feb-2001  
 C:Accession: T49064  
 R:Benes, V.; Wurmbach, E.; Drzonek, H.; Ansorek, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.; submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z25014  
 A:Accession: T49064  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-499 <BEN>  
 A:Cross-references: EMBL:AL163972; GSPDB:GN00061; ATSP:TSP19.280  
 A:Experimental source: cultivar Columbia; BAC clone TSP19  
 C:Genetics:  
 A:Gene: ATSP-TSP19.280  
 A:Map position: 3  
 A:Introns: 153/2  
 C:Superfamily: Candida cytochrome P450 homology  
 F:295-467/Domain: cytochrome P450 homology <P45>

Query Match 28.6%; Score 879.5; DB 2; Length 499;  
 Best Local Similarity 38.8%; Pred. No. 2.8e-60;  
 Matches 205; Conservative 96; Mismatches 176; Indels 51; Gaps 13;

QY 27 IALL--VLSWILVQWLSLRKQKGRSPVIGATVEQLARNYHRMDLWLVGYLSRHRTVT 84  
 DB 7 ISLLCPPIVIFFFPKNS--SSEFGKSPYIVGSLPGLVNNRHRFLDWTVELTSCRPTQT 65  
 QY 85 VDM---PFTSYIADPNVNEHVLKNTFTNPKGIVRSYMDVLLGGIENAGELWRKQ 141  
 DB 66 AIFRRPKLQFVMTANPANVEYMLKTFESFPKGERISILEDFLRGIFNSDGEMWKKQ 125  
 QY 142 RKTAPEFASKNLKDF--SAIVREYSLSKLSLSOASKAGKVDMDQSLYMRMTLDSICKV 200  
 DB 126 RKTAPEFASKNLKDF--SAIVREYSLSKLSLSOASKAGKVDMDQSLYMRMTLDSICKV 185  
 QY 201 GFGVEIGTSLPDLNEN--SFAQAFDAANIIITLRF--IDPLWRIKRFPHVGSALLAQSI 256  
 DB 186 AFNVDSACLDDGAGVNFQAFETAATIIISQRFQSVISYKWKIKKLNIGSERVLRESI 245  
 QY 257 KLVDDEFTYSVIRRRKAEIVVRASGKQEKMDILSLFELGEAGDGGGFGDDKSLRDV 316  
 DB 246 MIVHKFADEIVRNR---IE---QGVSDHEDLRSFISKEE-----MNSPEILRDI 291  
 QY 317 VLNPIVAGRTTATTLNFTHMASHPDVAPKRLRELCAPAEARABEGVTLVLCGGADA 376  
 DB 292 VISFILAGRTTSSALLSNFVLLSMHPEVKDKILQELNSIRERTGK----- 337  
 QY 377 DDKAFAARVAQFAGLLTYDSLGKLVYLHACVETLRLYPVQDPKPGILEDDVLPDGTKV 436  
 DB 338 -----RIGEVYG---PEDLKMNYLHAAITSLRLYPPVPVDTMSCAEDNVLDPDGTFI 387  
 QY 437 RAGGWVTIVPSYSGRMENYNGWPDAAFRPERWINE--DCAFRNAPSPFKETAFAQAPRCLG 495  
 DB 388 GKDWGISYINAYANGMESINGKDCDRFDPERIMDETNGFGFGENPFPKPFPAFHAGRMCLG 447  
 QY 496 KDSAYLQMK--MALAILFRYSRLLEGHVPQVYRMMTILSMHGLKVRV 542  
 DB 448 KEMAYIQMKSIVAALIERFVVEVPKGRPERPILMSVTLIRIGLNVRV 495

RESULT 9  
 JC5965  
 Cytochrome P450 CYP86A1 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 28-Jul-2000  
 C:Accession: JC5965  
 R:Benveniste, I.; Tijer, N.; Adas, F.; Philipps, G.; Salauen, J.P.; Durst, F.; Biochem. Biophys. Res. Commun. 243, 688-693, 1998  
 A:Title: CYP86A1 from Arabidopsis thaliana encodes a cytochrome P450-dependent fatty acid  
 A:Reference number: JC5965; MUID:98162595; PMID:9500987  
 A:Accession: JC5965

A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-513 <BEN>  
 A:Cross-references: EMBL:X90458  
 C:Superfamily: Candida cytochrome P450 homology  
 F:300-478/Domain: cytochrome P450 homology <P45>

Query Match 28.5%; Score 876.5; DB 2; Length 513;  
 Best Local Similarity 39.6%; Pred. No. 5e-60;  
 Matches 210; Conservative 83; Mismatches 182; Indels 55; Gaps 13;

QY 27 IALLLVLSWILVQWLSLRKQKGRSPVIGATVEQLARNYHRMDLWLVGYL-----SRHRT 82  
 DB 13 VAAVLSVALWFY---FLSRLTGPKVLPVGSPLPYLIANRSRIHDIADNLRTATGTYQT 69  
 QY 83 VTVDMPFTS-----YTYIADPNVNEHVLKNTFTNPKGIVRSYMDVLLGGIENAGEL 137  
 DB 70 CTWTFPFAKAQGFYTVTCHPKXNVEHILKTRFDNYPKGPMPWRAAFHDLGQGFNSDGT 129  
 QY 138 WRKQRTASFEFASKNLKDFSA--IVFREYSLKLSLSOASKAGKVDMDQSLYMRMTLDS 196  
 DB 130 WLMQRKTAALFETTTTLROAMRWVNGTIXKRLWLLDRVQNNKFPVDLQDLFLRLTFDN 189  
 QY 197 ICKVGFGEIGTSLPDLNEN--SFAQAFDAAN--IIITLRFIDPLWRIKRFPHVGSALLAQ 254  
 DB 190 ICLIFGKDPKPEILSLDLPNPFSAVDATETLKLTYTGFLWRIQKAMGIGSEDKLKK 249  
 QY 255 STKLVDDEFTYSVIRRRKAEIVVRASGKQEKMDILSLFELGEAGDGGGFGDDKSLR 314  
 DB 250 SLEVVETYNDAIDARK-----NSPDDLLSRFL---KKRDVNGVNLPTDVLQ 294  
 QY 315 DVVINEVIAGRTTATTLNFTHMASHPDVAPKRLRELCAPAEARABEGVTLVLCGGA 374  
 DB 295 RIALLNVLAGRTTSSVALLSNFVLLSMHPEVKDKILQELNSIRERTGK-----SMVLKEIR 341  
 QY 375 DADKFAARVAQFAGLLTYDSLGKLVYLHACVETLRLYPVQDPKPGILEDDVLPDGT 434  
 DB 342 GNDQEKWTEPLEF-----DEADRLVYLKAAAEITLRLYPSVPQDFKYVVEHDLVDPGT 395  
 QY 435 KVRAGGWVTIVPSYSGRMENYNGWPDAAFRPERWINE--DCAFRNAPSPFKETAFAQAPR 491  
 DB 396 FVPRGSTVTIVSYISIGRMKTINGEDCLFPRPRLMTADGE--RFETPKDGYKVFVAFNAGR 454  
 QY 492 ICLGKSAYLQMK--MALAILFRYSRLLEGHVPQVYRMMTILSMHGLKV 540  
 DB 455 TCLGKDLAYNQMKSVASVALLIRYVFP--VEGHRVEQKNSLTLFKNGLRV 503

RESULT 10  
 B86379  
 Protein F21J9.20 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: B86379  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: B86379  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-522 <STO>  
 A:Cross-references: GB:AB005172; NID:g9743340; PIDN:AAF97964.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F21J9.20  
 A:Map position: 1

C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

Query Match 27.2%; Score 837; DB 2; Length 522;

Best Local Similarity 36.2%; Pred. No. 6e-57; Indels 52; Gaps 11; Matches 192; Conservative 83; Mismatches 204; Indels 52; Gaps 11;

QY 27 IALLVVLVSLVQWLSLRKQKPRSPVIGATVEQLNRYHRMHDMLV-----GYLSRH 80  
 DB 20 VALAVGLFLLSYLREKLVSKGPPVMPVLGIIPMLALNKHDLFTWCTRCWRSGGFHY 79  
 QY 81 RTVTVDMPFTSYTYIADPNVNHVHLKTNFTNPKGIVYRSYMDVLLGDGFIADGELWRK 140  
 DB 80 RGIWFGAYGIMT--ADPANVHILKTNFTNPKGIVYRSYMDVLLGDGFIADGELWRK 137  
 QY 141 ORKTASFEFASKNLRDFAIVFRE-YSLKSLGSLQSKAGKVVMDQELVYRMFLDSICK 199  
 DB 138 ERRVAKTEHSSRLEHTFTMRDLVDQKLVPLNENLSTKRVFEDLQDLRLFRFFDNICI 197  
 QY 200 VGFGEIGTSLDPLPENSFAQADAAANIITLRFIDP--LMRIKRFHVHVSSEALLAOSIK 257  
 DB 198 SAFGVYPSLSTGLPEIPFAKAFEDATRYTLARELIPFVWPKMRFGLIGYERKLNNVR 257  
 QY 258 LVDEFTSVYRRKAEIVVEVRSQKQKHK-----DILSRFIEL---GEAGDDGGFGD 309  
 DB 258 IVHAFANKTVERR-----NMRKLGINDYADLLSLMOREVEYKESDTRTGNFYFS 308  
 QY 310 DKSIRDVVLNFIAGRDTTATTLGFWTHVMASHPDVAEKLRLRELCAPFAERAREEGVTLV 369  
 DB 309 DKYREFCTSIAGRDTTVALVWFVWLVQKHVEKRIILREIKRKLITQF----- 363  
 QY 370 LCGGADADKAFARVAQFAGLLTYDSLGKLVYLHACVTEITLYPVPQDPKGLLEDDV 429  
 DB 364 -----TED-----QFEA-----EDFRENVLQAALTSRLYPSVPMEMKQALEDDV 405  
 QY 430 LPDGTQVAGKVTVPYSGRMENYNGPDAAFPSPERWINEGDAFRNAPFKFTAFQAG 489  
 DB 406 LPDGTQVAGKVTVPYSGRMENYNGPDAAFPSPERWINEGDAFRNAPFKFTAFQAG 464  
 QY 490 PRICLGKDSAYLQMKWALILFRFYSFRLLGHPVQVYRMHTILSMAGHLKV 540  
 DB 465 PRLCVGKFAFQMKWAAAILMRISVKVQVQCEIVPKLITLYLMKNGMNV 515  
 RESULT 11  
 T46196  
 Cytochrome P450-like protein - Arabidopsis thaliana  
 N:Alternate names: protein T8P19.30  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 16-Feb-2001  
 C:Accession: T46196  
 R:Choi, S. N.; Robert, C.; Brothier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa  
 Submitted to the Protein Sequence Database, December 1999  
 A:Reference number: 223008  
 A:Accession: T46196  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-506 <CHO>  
 A:Cross-references: EMBL:AL133315  
 A:Experimental source: Cultivar Columbia; BAC clone T8P19  
 C:Genetics:  
 A:Map position: 3  
 A:Note: T8P19.30  
 C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
 P:300-469/Domain: cytochrome P450 homology <P45>  
 Query Match 26.7%; Score 823; DB 2; Length 506;  
 Best Local Similarity 36.1%; Pred. No. 6.9e-56; Indels 62; Gaps 14; Matches 193; Conservative 92; Mismatches 188; Indels 62; Gaps 14;  
 QY 27 IALLVVLVSLVQWLSLRKQK-----GPRSPVIGATVEQLNRYHRMHDMLVGY--LSRH 80  
 DB 9 ILAFLITIFLSSSTKVKQENTTYGPPSPILGSLISFNKRNHLLQWYTLRLSPS 68

QY 81 RTVTVDMPFTSYTYIADPNVNHVHLKTNFTNPKGIVYRSYMDVLLGDGFIADGELWR 139  
 DB 69 QUILVPLGNRRITITTNPLNVEYILKTNFFPKGKPTDILGLLGGGIFNVDSHS 128  
 QY 140 KORKTASFEFASKNLRDFAIVFRE-EYSLKSLGSLQSKAGKVVMDQELVYRMHTILDSIC 198  
 DB 129 SQKSLASHSEFSRSLRSFAFVLEKOBVENRLVPLSTAADVGTTVDLQDVLKRFAPDVVC 188  
 QY 199 KVGFGVEIGTSLDPLPENSFAQADAAANIITLRFIDP--LMRIKRFHVHVSSEALLAOS 255  
 DB 189 KVSGLWDPCDCLDITREVNPLVEAFDAEISARRATEPIYAVMKTKRVNLVSGSEKRLRA 248  
 QY 256 IKLVDEFTSVYRRKAEIVVEVRSQKQKHKHILSRFIELCEAGDDGGFGDCKSLRD 315  
 DB 249 IRVTHVLSVSEIVRAKKSL-EI---GTGAQKQDILSRFLAAGHGE-----AVRD 295  
 QY 316 VLNFIAGRDTTATTLGFWTHVMASHPDVAEKLRLRELCAPFAERAREEGVTLVLCGGAD 375  
 DB 296 MVISFINAGRDTTSAATMLFWLLTENDDVERKI-----LEEVDPLVSLG--- 340  
 QY 376 ADDKAFARVAQFAGLLTYDSLGKLVYLHACVTEITLYPVPQDPKGLLEDDVLPDGT 435  
 DB 341 -----LGFDLKEAMAYTKACLEAMRLYPPVSWDSKHAANDDVLDPDGT 384  
 QY 436 VRAGGMVYVYSGRMENYNGPDAAFPSPERWI-NEDGAFRNA---SPFKFTAFQAGP 490  
 DB 385 VKRGDKVTYTPYSGRMENYNGPDAAFPSPERWI-NEDGAFRNA---SPFKFTAFQAGP 444  
 QY 491 RICLGKDSAYLQMKWALILFRFYSFRLLGHPVQVYRMHTILSMAGHLKVRSR 544  
 DB 445 RVCVGKFAFQMKWAAAILMRISVKVQVQCEIVPKLITLYLMKNGMNV 498  
 RESULT 12  
 G86265  
 F3P19.17 protein - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
 C:Accession: G86265  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Corway, A.R.; Creasy, T.H.; Dewar, K.  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: G86265  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-529 <STO>  
 A:Cross-references: GB:AE005172; NID:G4850397; PIDN:AD31067.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
 Query Match 25.6%; Score 788; DB 2; Length 529;  
 Best Local Similarity 34.2%; Pred. No. 3.8e-53; Indels 64; Gaps 13; Matches 187; Conservative 102; Mismatches 193; Indels 64; Gaps 13;  
 QY 24 HKYIALLVLSWI-----LVQWLSLRKQKPRSPVIGATVEQLNRYHRMHDMLVGLS 78  
 DB 8 YNHLISLFDLLLSLLGLFVFCCLREKLTNRKGMPLPVFGITLFFHINDVYWGTVRSUK 67  
 QY 79 R-----HRTVTVDMPFTSYTYIADPNVNHVHLKTNFTNPKGIVYRSYMDVLLGDGFI 132  
 DB 68 KCRGFLYRGWLDGSGYAVTCV--PANVEYMLKTNFNKFNPKTFKSRFNDLLEGIEN 125  
 QY 133 ADGELWRKQKRTASFEFASKNLRDFAIVFREYSL-----KLSGLISQSKAGKVV 183

Db 126 ADDSNKQRIIIITEMHSTG-----FVHSGFQTTQHLVRKLLKVMBSFAKSOBAF 177  
 QY 184 DMQELYMRTLDISCKVGFGEIGTSLSPDIPNSFAQAADANIIITLRFIDP--LWRIK 241  
 Db 178 DLQDVLRLTFTDIIICLAGLGADPEYLAVDLPQVPAKAFEAETSTLFRFPIPPWPKM 237  
 QY 242 RFFHVGSEALLAQSIKLVDFTYSVIRRKAEIVVRASGKQEKM--RHDILSRFIEL-G 298  
 Db 238 RLDTGYEKGLRIANGVGHGFDVKMIVDRICELKE-----EETLDRSDVLRIQIES 291  
 QY 299 EAGDGGGFGDDKSLRDVVLNFIAGRTTATLWNFTMHAMSHPDVAEKLRELCAFE 357  
 Db 292 HKRENEIDPSTIRFRQFCTSPILAGRTDSSVALSWFCWVIOKHPEVENKI---ICEIRE 348  
 QY 358 ABRAREEGVTVLVCGGADDAKFAARVAQFAGLTYSGLKVLVYHACVTETLRLYPAV 417  
 Db 349 ILRQGRDSTP-----SKNSLFTVVKELNNVYLOALSTLRLFPPI 390  
 QY 418 PODPKGILEDDVLPDGTCKVRAGMVTYVPSYGRMEYNWGPDAASFRRPRWINEGAFRN 477  
 Db 391 PHEMKQAIEDDVLDPGTFVRKGRSVVFSIYANGRMESIWGKDCETFRPRPWI-QAGKFS 449  
 QY 478 ASPFKFTAFQAGRICLGKDSAYLQMKMALILFRYSFRLLECHPVQVRMTILSMANG 537  
 Db 450 DDQFKVWFENAGPRLCIGKTFAYLQMKMIAASVLLFRYSIKVQDQHVIAPRVTNLYMYKG 509  
 QY 538 LKVRVS 543  
 Db 510 LKVTIT 515

RESULT 13  
 T08014  
 cytochrome P450 CYP94A1 - spring vetch  
 C:Species: *Vicia sativa* (spring vetch, tare)  
 C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 16-Feb-2001  
 C:Accession: T08014  
 R:Tijer, N.; Helvig, C.; Pinot, F.; le Bouquin, R.; Lesot, A.; Durst, F.; Salaun, J.P.;  
 Biochem. J. 332, 583-589, 1998  
 A:Title: Functional expression in yeast and characterization of a clofibrate-inducible P  
 A:Reference number: Z16287; MUID:98264856; PMID:9601090  
 A:Accession: T08014  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-514 <TJJ>  
 A:Cross-references: EMBL:AF030260; NID:G4204094; PIDN:AAD10204.1; PID:G4204095  
 C:Gene: vagn11  
 C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
 F:312-480/Domain: cytochrome P450 homology <P45>

Query Match 25.5%; Score 785.5; DB 2; Length 514;  
 Best Local Similarity 35.0%; Pred. No. 5.8e-53;  
 Matches 165; Conservative 95; Mismatches 154; Indels 57; Gaps 8;

QY 50 PRSPVLTGATVEQURNYHRMDML--VGYLSRHRTVTVDMPFTSYTYI-ADPNNVHVXK 106  
 Db 48 PKSYPLGSLVSPKXNLRHRLQWLSDIVQSPATFQDGLGKRIITGNPSTVQHIK 107  
 QY 107 TNFTNPKGIYVRMYDVLGDTGFNADGELWRKQRTASPEFASKMLRDF-SALIVREY 165  
 Db 108 NQFSNYQKGTFTTSLDFTLGTGINTNGPNWKTQROVASHEFNKTGIRNVEHIVDTL 167  
 QY 166 SLKLSILSQASKAGKVVDMQELYWRMTLDSICKVGFGEIGTSLSPDIPNSFAQAADAN 225  
 Db 168 TNRILPILTSQTNNILDFQDILQRFTDNCIMAFQYDEYITPSTNRKFAEVEDA 227  
 QY 226 NIITLRFIDPL--WRKRFRHVGSEALLAQSIKLVDFTYSVIRRKAEIVVRASGK 282  
 Db 228 TEISKRRLFLPLPIWIKKYFNIGSEKELAEVTEVRSFAKLVREKRELEE-----K 282  
 QY 283 QEKMKHILSRFIELGEAGDGGGFGDDKSLRDVVLNFIAGRTTATLWNFTMHAMSH 342

Db 283 SSLETDMLSRFL-----SSCHSDEDFVADIVISFILLAGKDTTSAALTWFELLWKN 334  
 QY 343 PDVAEKLRELCAFEABRAREEGVTVLVCGGADDAKFAARVAQFAGLTYSGLKVLVY 402  
 Db 335 PRVEEIVNEL-----SKGBELMAYDEVKEMVY 362  
 QY 403 LHACVTETLRLYPVPOPKGILEDDVLPDGTCKVRAGMVTYVPSYGRMEYNWGPDAAS 462  
 Db 363 THAALSMSRLYPPVPMDSKEAVNDVLPDGNVVKVGTIVTYHYVYANGRMKSLGDDWAE 422  
 QY 463 FRPRWINEED-----GAFRNASPFKFTAFQAGRICLGKDSAYLQMKMALA 508  
 Db 423 FRPRWLEKDEWNGKWFVGRDSYSYPVQAGPRVCLGKEMAFMQMKRIVA 473

RESULT 14  
 F86265  
 hypothetical protein F3F19.16 - Arabidopsis thaliana  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C:Accession: F86265  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ausen, N.P.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: F86265  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-519 <STO>  
 A:Cross-references: GB:AE005172; NID:G4850398; PIDN:AAD31068.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

Query Match 25.3%; Score 780; DB 2; Length 519;  
 Best Local Similarity 34.1%; Pred. No. 1.6e-52;  
 Matches 184; Conservative 91; Mismatches 187; Indels 78; Gaps 12;

QY 27 IALLVVLVSWILVQWLSRKQKQPRSPVIGATVEQLRNYHRMDMLVGVLSR-----H 80  
 Db 23 VSLALLGLFVCCVREKVKLGTITVPVFGITPEFFHRNDVYGVATRCCKKCGRTFLY 82  
 QY 81 RTVTVDMPFTSYTYIADPVNVHVLKTNFTNPKGIYVRMYDVLGDTGFNADGELWRK 140  
 Db 83 NGIWLGGSYGAVTCV--PANVEYMLKTNFKNPKGAFKFERFNDLLEDGIFNADAESWKE 140  
 QY 141 ORKTASPEFASKNLRDFAIVFREYSL-----KLGIILSQASKAGKVVDMQELYMR 191  
 Db 141 ORIIITEMISTR-----FVHSGFQTTQHLVRKLLKVMBSFAKSOBAF 192  
 QY 192 MTLDSCICKVGFGEIGTSLSPDIPNSFAQAADANIIITLRFIDP--LWRIKFEHVGSE 249  
 Db 193 LTFDNICIAGLDGDDPGTLDSDPLVPPAQAFAEATSTMFPMIPPPKPLKFFDGYE 252  
 QY 250 ALLAQSTKLVDFTYSVIRRKAEIVVRASGKQEKMHDLISRFIELGEAGDGGGFGD 309  
 Db 253 KGLRKAV-----DVSMSLSLSTR---WLIVVSASSKKKQSHKTT-----D 288  
 QY 310 DKS-----LRDVVLNFIAGRTTATLWNFTMHAMSHPDVAEKLRELCAFEABRARE 363  
 Db 289 EKDPSTIKFRQFCTSPILAGRTSSVALTWFFWVIOKHPEVENKIIREI---SEILRQ 344  
 QY 364 EGVTVLVCGGADDAKFAARVAQFAGLTYSGLKVLVYHACVTETLRLYPVPOPKG 423  
 Db 345 RG-----DSFTSKNESILFTVKELNDMVYLQAAALSETMRLYPPPIPMEMKQ 388

Search completed: March 2, 2004, 15:29:43  
Job time : 22 secs

NO

NO



181 QY AGCAATGAGAGATCTATTACAGACTTACACTGATGCCGCCAACAATTAACCTTTAGA 240  
181 Db AGCAATGAGAGATCTATTACAGACTTACACTGATGCCGCCAACAATTAACCTTTAGA 240  
241 QY TTTTGATATGTCACCTCCTACTTATCTCTGTTGGGCAACTTCCAAATAGGCTCATGTT 300  
241 Db TTTTGATATGTCACCTCCTACTTATCTCTGTTGGGCAACTTCCAAATAGGCTCATGTT 300  
301 QY AATCAATGATTAAGTATTAATCAGCAAAATATTCTTCTTTGTTTGTAGCAATTAATAATG 360  
301 Db AATCAATGATTAAGTATTAATCAGCAAAATATTCTTCTTTGTTTGTAGCAATTAATAATG 360  
361 QY GGTGAGACGGATTAATAATCATCCATGAGAGCTTTATCTTCAATGCTCTCTTGAATTTGG 420  
361 Db GGTGAGACGGATTAATAATCATCCATGAGAGCTTTATCTTCAATGCTCTCTTGAATTTGG 420  
421 QY TTTTCAGATCATCTTTTCAAGTGTTCACAGAAATTTCTCAGTTTGGTCCCAATTAATTTTG 480  
421 Db TTTTCAGATCATCTTTTCAAGTGTTCACAGAAATTTCTCAGTTTGGTCCCAATTAATTTTG 480  
481 QY AAGTGAAGTTTCTTAATTTTCAATATGCTTCTTCTTTCTAGACTAGCAACTGCATGA 540  
481 Db AAGTGAAGTTTCTTAATTTTCAATATGCTTCTTCTTTCTAGACTAGCAACTGCATGA 540  
541 QY CTTTTCACCTTTGGGTTCACAAAATGACTCAAGAAATAAATAATTTCACTTTTGGTTTACA 600  
541 Db CTTTTCACCTTTGGGTTCACAAAATGACTCAAGAAATAAATAATTTCACTTTTGGTTTACA 600  
601 QY AATTCCTCTTCAAGTGTACTTTTCACTTCTGAACTGTATAGGAACAAGGAATGGCT 660  
601 Db AATTCCTCTTCAAGTGTACTTTTCACTTCTGAACTGTATAGGAACAAGGAATGGCT 660  
661 QY CAGTTTTTAAGGAACAATGATACAGATTTCAATTTCAAGAACTCTTTCTGGTTGGTTGAGTT 720  
661 Db CAGTTTTTAAGGAACAATGATACAGATTTCAATTTCAAGAACTCTTTCTGGTTGGTTGAGTT 720  
721 QY CAGACTTTTTGTACCAAGCTGATGGATFACAATATCTTGTTCACAAAGTCTGATAACAGAA 780  
721 Db CAGACTTTTTGTACCAAGCTGATGGATFACAATATCTTGTTCACAAAGTCTGATAACAGAA 780  
781 QY ACTGGCAACTCTTAATTTGATAATAAAGAAATAAATAACAGATATCAGATATCTCATTTTC 840  
781 Db ACTGGCAACTCTTAATTTGATAATAAAGAAATAAATAACAGATATCAGATATCTCATTTTC 840  
841 QY TTGGTTGGCAGATCAAAAAGGAACAACAAGGCTAAGCTCTCTACTTGTTCGGGAGTTA 900  
841 Db TTGGTTGGCAGATCAAAAAGGAACAACAAGGCTAAGCTCTCTACTTGTTCGGGAGTTA 900  
901 QY GGTGAGGACACCATATGAAATGAAGAAATCTTAATTTGGGGTCAACCAAGATTTGCTC 960  
901 Db GGTGAGGACACCATATGAAATGAAGAAATCTTAATTTGGGGTCAACCAAGATTTGCTC 960  
961 QY TCTCAGGTTGGGGGTCCTTAAGTTTGGTAGTACCAATACCCAAATATATCACTTAACAA 1020  
961 Db TCTCAGGTTGGGGGTCCTTAAGTTTGGTAGTACCAATACCCAAATATATCACTTAACAA 1020  
1021 QY ACCCAATCCATGCTACATACATACATAGCAATCCATCTGATGAGAGCTTCCATCA 1080  
1021 Db ACCCAATCCATGCTACATACATACATAGCAATCCATCTGATGAGAGCTTCCATCA 1080  
1081 QY AGAGCACCATGAGAGAGCTCACATACGCGCGGACGCCATCCCAATCTTCCCACTAG 1140  
1081 Db AGAGCACCATGAGAGAGCTCACATACGCGCGGACGCCATCCCAATCTTCCCACTAG 1140  
1141 QY CAGGCTCTCAAGTATCATCGGCTCTCTCTGTTGTCTCTCTCATGATCTCTGTTCCAGA 1200  
1141 Db CAGGCTCTCAAGTATCATCGGCTCTCTCTGTTGTCTCTCTCATGATCTCTGTTCCAGA 1200  
1201 QY GGTGAGGCTTGAAGAGCAGAAAGGCCCGAGATCATATGCGGCTCATCGGTCAACGGTGG 1260  
1201 Db GGTGAGGCTTGAAGAGCAGAAAGGCCCGAGATCATATGCGGCTCATCGGTCAACGGTGG 1260  
1261 QY AGCAGCTGAGGAACTACCAACGGATGACGACTGGCTTGTGGGTACCTGTACGGGCACA 1320

1261 Db AGCAGCTGAGGAACTACCAACGGATGACGACTGGCTTGTGGGTACCTGTACGGGCACA 1320  
1321 QY GGACAGTACCGTTCGACATGCGGTTCACTTCTTACCTACCTACCTGACCGGTTGAATG 1380  
1321 Db GGACAGTACCGTTCGACATGCGGTTCACTTCTTACCTACCTACCTGACCGGTTGAATG 1380  
1381 QY TCGAGCATGCTCTCAAGACTAATTTCAACAATTAACCAAGTAAATGACCTGAATCAAC 1440  
1381 Db TCGAGCATGCTCTCAAGACTAATTTCAACAATTAACCAAGTAAATGACCTGAATCAAC 1440  
1441 QY TGATGCTTCACTTCCGAAATCAGAGCTGAAAGCTGAAATCGAATGCTGCTGAACACCGTG 1500  
1441 Db TGATGCTTCACTTCCGAAATCAGAGCTGAAAGCTGAAATCGAATGCTGCTGAACACCGTG 1500  
1501 QY TAGGAAATCGTGTACAGATCTTACATGGAACGCTCTCGGTGACGGCATCTTCAACGCG 1560  
1501 Db TAGGAAATCGTGTACAGATCTTACATGGAACGCTCTCGGTGACGGCATCTTCAACGCG 1560  
1561 QY GACGGGAGCTGTGAGGAGCAGAGGAGAGCGGAGGTTTCGAGTTGCGCTCCAAGAAC 1620  
1561 Db GACGGGAGCTGTGAGGAGCAGAGGAGAGCGGAGGTTTCGAGTTGCGCTCCAAGAAC 1620  
1621 QY CTGAGGATTTTCAAGCGCATTTCTGTTTCAAGAGTAGTACTTCTTGAAGCTGTGCGGTATCTG 1680  
1621 Db CTGAGGATTTTCAAGCGCATTTCTGTTTCAAGAGTAGTACTTCTTGAAGCTGTGCGGTATCTG 1680  
1681 QY AGCCAGGATCAAGGAGGCAAAAGTTTGTGGAATGAGTATGAGTATGATCTCTCTTTC 1740  
1681 Db AGCCAGGATCAAGGAGGCAAAAGTTTGTGGAATGAGTATGAGTATGATCTCTCTTTC 1740  
1741 QY CATTTGCCAATGAGCATTTCAACCTGAGACAGAGACTACCTTTCGATTTCAAGAACT 1800  
1741 Db CATTTGCCAATGAGCATTTCAACCTGAGACAGAGACTACCTTTCGATTTCAAGAACT 1800  
1801 QY TTACATGAGGATGACGCTGGAATCTTCAATCTGCAAGGTTTGGGTTTGGGGTCAAGATCGGCAC 1860  
1801 Db TTACATGAGGATGACGCTGGAATCTTCAATCTGCAAGGTTTGGGTTTGGGGTCAAGATCGGCAC 1860  
1861 QY GCTGTCCGCGGATCTCCCGAGAACAGACTTTCGCGAGGGGTTTCTGATGCCGCCAATCAT 1920  
1861 Db GCTGTCCGCGGATCTCCCGAGAACAGACTTTCGCGAGGGGTTTCTGATGCCGCCAATCAT 1920  
1921 QY GGTCAAGCTGCGGTTTCACTGACCGCTGCGGATCAAGAGGTTTCTTCCAGCTCGGGTC 1980  
1921 Db GGTCAAGCTGCGGTTTCACTGACCGCTGCGGATCAAGAGGTTTCTTCCAGCTCGGGTC 1980  
1981 QY AGAGGCTCTCTAGCGCAGAGCATCAAGCTCGTGGACGAGTTTCACTACAGCGTGTATCCG 2040  
1981 Db AGAGGCTCTCTAGCGCAGAGCATCAAGCTCGTGGACGAGTTTCACTACAGCGTGTATCCG 2040  
2041 QY CCGGAGAAAGCCGAGATCGTTCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100  
2041 Db CCGGAGAAAGCCGAGATCGTTCGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100  
2101 QY ACATGACTGTTTTCGATTTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTT 2160  
2101 Db ACATGACTGTTTTCGATTTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTT 2160  
2161 QY ATTATATATCCGTTGATTTTGTGAGGACAAATTAATAATGAGGAGATGAAGCAGCATCC 2220  
2161 Db ATTATATATCCGTTGATTTTGTGAGGACAAATTAATAATGAGGAGATGAAGCAGCATCC 2220  
2221 QY TGTCAAGGTTTCATGAGCTAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2280  
2221 Db TGTCAAGGTTTCATGAGCTAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2280  
2281 QY AGAGCTCTCCGCGAGCTGCTCAATTTCTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2340  
2281 Db AGAGCTCTCCGCGAGCTGCTCAATTTCTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2340  
2341 QY CGTGTCTGCTGTTTCAAGCAGATGCGGCTTCCCAACCGGAGCTGCGGCGGCGGCGGCGGCGG 2400  
2341 Db CGTGTCTGCTGTTTCAAGCAGATGCGGCTTCCCAACCGGAGCTGCGGCGGCGGCGGCGGCGG 2400

Db 2341 CGCTGTGCTGGTTACGCAATGCCCATGTCCCAACCGGACGTGGCCGAGAAAGCTGCGCC 2400  
Qy 2401 GCGAGCTGTGCGGCTTCGAGCGGAGCGCGCGCGGAGGAGGCGTTCGCGCTCGTGCCT 2460  
Db 2401 GCGAGCTGTGCGGCTTCGAGCGGAGCGCGCGCGGAGGAGGCGTTCGCGCTCGTGCCT 2460  
Qy 2461 GCGCGGCGCTGAGCGCGAGCAGAAAGGCTTCGCGCGCGCGCGGAGTTCGCGGCGC 2520  
Db 2461 GCGCGGCGCTGAGCGCGAGCAGAAAGGCTTCGCGCGCGCGCGGAGTTCGCGGCGC 2520  
Qy 2521 TCCTCACTACGACGCTCGGCAAGCTGTCTTACCTCCAGCGCTGGTCAACGAGACGC 2580  
Db 2521 TCCTCACTACGACGCTCGGCAAGCTGTCTTACCTCCAGCGCTGGTCAACGAGACGC 2580  
Qy 2581 TCCGCTGTATCCCGCGCGCTCCCTCAGGTGAGCGCGCGCGAGCAACGCGACTCCGTTCCAGA 2640  
Db 2581 TCCGCTGTATCCCGCGCGCTCCCTCAGGTGAGCGCGCGCGAGCAACGCGACTCCGTTCCAGA 2640  
Qy 2641 GCACAGATGAGTGGACCTGAATGCAATGCATGCACTTGGCGCGCGCGAGAC 2700  
Db 2641 GCACAGATGAGTGGACCTGAATGCAATGCATGCACTTGGCGCGCGCGAGAC 2700  
Qy 2701 CCCAAGGCGATCTGGAGGACGAGCTGTGCGGACGCGAGCAAGGTGAGGCGCGCGG 2760  
Db 2701 CCCAAGGCGATCTGGAGGACGAGCTGTGCGGACGCGAGCAAGGTGAGGCGCGCGG 2760  
Qy 2761 ATGTTGAGCTACGTCCCTACTCGATGAGGCGGATGAGAGTCAAACTGGGCGCGCGC 2820  
Db 2761 ATGTTGAGCTACGTCCCTACTCGATGAGGCGGATGAGAGTCAAACTGGGCGCGCGC 2820  
Qy 2821 GCGAGTCTCCGCGCGGAGCGGTGATCAACAGAGTGGCGGTTCCGCAACGCGTCCGCG 2880  
Db 2821 GCGAGTCTCCGCGCGGAGCGGTGATCAACAGAGTGGCGGTTCCGCAACGCGTCCGCG 2880  
Qy 2881 TTCAAGTTTCAAGGCTTCAGCGCGGCGCGGAGCTGCTGGCAAGGACTCGGCGTAC 2940  
Db 2881 TTCAAGTTTCAAGGCTTCAGCGCGGCGCGGAGCTGCTGGCAAGGACTCGGCGTAC 2940  
Qy 2941 CTGCAATGAAGATGCGCTGCGCATCTCTTGGCGCTTCAAGCTTCGCGCTGCTGAG 3000  
Db 2941 CTGCAATGAAGATGCGCTGCGCATCTCTTGGCGCTTCAAGCTTCGCGCTGCTGAG 3000  
Qy 3001 GGGCAACCGGTCAGTACCGCATGATGACCATCTCTCCATGGCGCAGGCTCAAGTTC 3060  
Db 3001 GGGCAACCGGTCAGTACCGCATGATGACCATCTCTCCATGGCGCAGGCTCAAGTTC 3060  
Qy 3061 CGCGTCTTAGGCGCTCTGATGTATGCGGATTTGGGATATCATCCCGCTTAATCCTTA 3120  
Db 3061 CGCGTCTTAGGCGCTCTGATGTATGCGGATTTGGGATATCATCCCGCTTAATCCTTA 3120  
Qy 3121 AAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180  
Db 3121 AAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180  
Qy 3181 CTTTAAATCTCGTGGCTTCGGAACACACATCACTAGTGTGTTGATCTACTCT 3240  
Db 3181 CTTTAAATCTCGTGGCTTCGGAACACACATCACTAGTGTGTTGATCTACTCT 3240  
Qy 3241 CTCAGTGAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300  
Db 3241 CTCAGTGAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300  
Qy 3301 CGGATCTTCCGCGGACCTTTGGAGACCATTAATGACAGGCGGTGTGAAAAAGGCTTC 3360  
Db 3301 CGGATCTTCCGCGGACCTTTGGAGACCATTAATGACAGGCGGTGTGAAAAAGGCTTC 3360  
Qy 3361 TTCGCGGCAAGTTCGAGTCTGAGTCTTGGCGCTTTCGAGCAGAAAAAGGTTTCG 3420  
Db 3361 TTCGCGGCAAGTTCGAGTCTTGGCGCTTTCGAGCAGAAAAAGGTTTCG 3420  
Qy 3421 AAGGATCTGAAACCGGAAATGGCTTCGGAATATGCTCGCATCGGCGCGGCGC 3480  
Db 3421 AAGGATCTGAAACCGGAAATGGCTTCGGAATATGCTCGCATCGGCGCGGCGC 3480

Qy 3481 GTCACTCGGATGACGACAGCCCAAGCAGTGAAGCGATCTTTTGAGTTTGG 3540  
Db 3481 GTCACTCGGATGACGACAGCCCAAGCAGTGAAGCGATCTTTTGAGTTTGG 3540  
Qy 3541 AGACACTCTCGGACCCCTCGGCGCTCCGCGAGCTCATCTTCGCTCTCTGTGTGTCG 3600  
Db 3541 AGACACTCTCGGACCCCTCGGCGCTCCGCGAGCTCATCTTCGCTCTCTGTGTGTCG 3600  
Qy 3601 TGGCGGACCGCGCGCGCGCTCGTTCGACCAATCCGCGCGCGCGCGGTTGCT 3660  
Db 3601 TGGCGGACCGCGCGCGCGCTCGTTCGACCAATCCGCGCGCGCGCGGTTGCT 3660  
Qy 3661 GTACAAACCTCTCATCCGCGCGCGCGTGAACAGCGCGCGCGCGCGCTATACAT 3720  
Db 3661 GTACAAACCTCTCATCCGCGCGCGCGTGAACAGCGCGCGCGCGCTATACAT 3720  
Qy 3721 CTATAATCATGTGTATTTTATTTTCAACCGGCTTAAACAAACCATATTTTATG 3780  
Db 3721 CTATAATCATGTGTATTTTATTTTCAACCGGCTTAAACAAACCATATTTTATG 3780  
Qy 3781 GTAAACAGCTTCAAAATTTGACAAATTTAAACAGGACCAACCGTAGCTAAACATAAG 3840  
Db 3781 GTAAACAGCTTCAAAATTTGACAAATTTAAACAGGACCAACCGTAGCTAAACATAAG 3840  
Qy 3841 AGAATGAGACACCAACCAAGGTTAGAGTGAATAAGCTGAGTAAACGACGAATTC 3897  
Db 3841 AGAATGAGACACCAACCAAGGTTAGAGTGAATAAGCTGAGTAAACGACGAATTC 3897

## RESULT 2

US-10-412-000-7  
; Sequence 7, Application US/10412000  
; Publication No. US2003018289A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; TITLE OF INVENTION: METHOD OF USING SAME  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 3897  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-412-000-7

Query Match 100.0%; Score 3897; DB 14; Length 3897;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3897; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAATTCGAAGCGAGCGCCCTTTAGCAGAGAGTGTTCGTGATGCGAGTTCGCGCGGAATGAGT 60  
Db 1 GAATTCGAAGCGAGCGCCCTTTAGCAGAGAGTGTTCGTGATGCGAGTTCGCGCGGAATGAGT 60  
Qy 61 GCGTCTGAGAGCAACGCTGAGGGTTCCAGGATGCAATGGCTATGCGAATCGGCTAG 120  
Db 61 GCGTCTGAGAGCAACGCTGAGGGTTCCAGGATGCAATGGCTATGCGAATCGGCTAG 120  
Qy 121 AGGTGAGGACCAAGTGTGTAGGATTTGGAGGGCAACCTATGGCAAGTTGGTGAAGAGGC 180  
Db 121 AGGTGAGGACCAAGTGTGTAGGATTTGGAGGGCAACCTATGGCAAGTTGGTGAAGAGGC 180  
Qy 181 AGCAATGAGAGATCTATTTCAGACTTACATGGATGCGGCCCAACAAATTCACCTTTAGA 240  
Db 181 AGCAATGAGAGATCTATTTCAGACTTACATGGATGCGGCCCAACAAATTCACCTTTAGA 240

QY 241 TTTTGTACTGTGACCTCACTACTTTATTTCTTGGTTGGCAACTTCCAATAGGCTCATGTT 300  
Db 241 TTTTGTACTGTGACCTCACTACTTTATTCCTTGGTTGGCAACTTCCAATAGGCTCATGTT 300  
QY 301 AATCAATGATAGTGAATATTCAGCAAAATATTCCTTGGTTGGTGTGACATTTAATAATG 360  
Db 301 AATCAATGATAGTGAATATTCAGCAAAATATTCCTTGGTTGGTGTGACATTTAATAATG 360  
QY 361 GGTGTGACGGATTAATAATCATCCATGAGAGCTTTATCTTCATGCTCTCTTCATTTTGG 420  
Db 361 GGTGTGACGGATTAATAATCATCCATGAGAGCTTTATCTTCATGCTCTCTTCATTTTGG 420  
QY 421 TTTTCAGATCATCTTTTCAGTGTTCACAGAAATTTTCTCAGTTTGGTCCATGTAATTTTGG 480  
Db 421 TTTTCAGATCATCTTTTCAGTGTTCACAGAAATTTTCTCAGTTTGGTCCATGTAATTTTGG 480  
QY 481 AAGTGAGGTTCTTAAATTTTCATATGCTTCCTTCTTCTTTCTAGACTAGCAACTGCATGA 540  
Db 481 AAGTGAGGTTCTTAAATTTTCATATGCTTCCTTCTTCTTTCTAGACTAGCAACTGCATGA 540  
QY 541 CTTTTCACCTTTGGTTCACAAAATTTGACTCACAGAAAACAAATTTCACTTTTGGTTTCACA 600  
Db 541 CTTTTCACCTTTGGTTCACAAAATTTGACTCACAGAAAACAAATTTCACTTTTGGTTTCACA 600  
QY 601 AATTTCCTCTTCAGGATGCTATTTTCACTTGAACCTGTCTATGTATAGGAACAAGGAATGGCT 660  
Db 601 AATTTCCTCTTCAGGATGCTATTTTCACTTGAACCTGTCTATGTATAGGAACAAGGAATGGCT 660  
QY 661 CAGTTTAAAGGAACAATGTACAGATTTCAATTTTCAGAACTCTTTCTGGTTGGTTGAGTTT 720  
Db 661 CAGTTTAAAGGAACAATGTACAGATTTCAATTTTCAGAACTCTTTCTGGTTGGTTGAGTTT 720  
QY 721 CAGACTTTTGTACCAAGCTGATGGATCACAATATCTTGTTCAAAAGCTCTGATAACAGAA 780  
Db 721 CAGACTTTTGTACCAAGCTGATGGATCACAATATCTTGTTCAAAAGCTCTGATAACAGAA 780  
QY 781 ACTGGCAACTCTTAATTTGATAATAAAGGAATAAATAACAGTATCAGATATCTCATTTTC 840  
Db 781 ACTGGCAACTCTTAATTTGATAATAAAGGAATAAATAACAGTATCAGATATCTCATTTTC 840  
QY 841 TTGGTTGGCAGATACAAAAGGAACAACAAGGCTAAGCTTCCTTCTTGTTCGGAGTTA 900  
Db 841 TTGGTTGGCAGATACAAAAGGAACAACAAGGCTAAGCTTCCTTCTTGTTCGGAGTTA 900  
QY 901 GGTCAAGGACACCATATGAATGAAGAATCTTAATTTGGGGTCAACCAAGATTTGTCTC 960  
Db 901 GGTCAAGGACACCATATGAATGAAGAATCTTAATTTGGGGTCAACCAAGATTTGTCTC 960  
QY 961 TCTCGAGGTTGGGGGTCCTTAAGTTGGTATAGCAATACCCAAATATACCTTAACAA 1020  
Db 961 TCTCGAGGTTGGGGGTCCTTAAGTTGGTATAGCAATACCCAAATATACCTTAACAA 1020  
QY 1021 ACCCAATCCATGCTACATACATACATAGCATCCATCTCTAGACTGGACCTTCATCA 1080  
Db 1021 ACCCAATCCATGCTACATACATACATAGCATCCATCTCTAGACTGGACCTTCATCA 1080  
QY 1081 AGAGCACCATGGAGGAGCTCACATCAGCCGGGAGCCCATCGCCATTCCTTCCCACTAG 1140  
Db 1081 AGAGCACCATGGAGGAGCTCACATCAGCCGGGAGCCCATCGCCATTCCTTCCCACTAG 1140  
QY 1141 CAGGGCCTCAAGATACATTCGGCTTCCTTCTGTTGTCTCTCATGGATCTTGTTCAGA 1200  
Db 1141 CAGGGCCTCAAGATACATTCGGCTTCCTTCTGTTGTCTCTCATGGATCTTGTTCAGA 1200  
QY 1201 GGTGTGAGCTGAGGAAGCAGAAAGCCGAGATCATGCGCCAGTCAATCGGTCAACGGTGG 1260  
Db 1201 GGTGTGAGCTGAGGAAGCAGAAAGCCGAGATCATGCGCCAGTCAATCGGTCAACGGTGG 1260  
QY 1261 AGCAGCTGAGGAACCTACCAACCGGATGCACTGCTGCTTGTGGGTACCTGTCAAGGCACA 1320  
Db 1261 AGCAGCTGAGGAACCTACCAACCGGATGCACTGCTGCTTGTGGGTACCTGTCAAGGCACA 1320

QY 1321 GGACAGTGACCGTTCGACATCCGTTTCATCTTCTACACCTACATCGCTGACCCGGTGAATG 1380  
Db 1321 GGACAGTGACCGTTCGACATCCGTTTCATCTTCTACACCTACATCGCTGACCCGGTGAATG 1380  
QY 1381 TCGAGCATGCTCTCAAGACTAACTTCAACAAATACCCAAAGTAAATGACCTGAACTCAC 1440  
Db 1381 TCGAGCATGCTCTCAAGACTAACTTCAACAAATACCCAAAGTAAATGACCTGAACTCAC 1440  
QY 1441 TGATGTTTCACTCTTCGGAATCAGAGCTGAAAGCTGAATCGAATGTCCTGAAACCGGTG 1500  
Db 1441 TGATGTTTCACTCTTCGGAATCAGAGCTGAAAGCTGAATCGAATGTCCTGAAACCGGTG 1500  
QY 1501 TAGGGAATCGTGTACAGATCTTACATGAGCTGCTCTCGGTGACGGCATCTTCAACGGC 1560  
Db 1501 TAGGGAATCGTGTACAGATCTTACATGAGCTGCTCTCGGTGACGGCATCTTCAACGGC 1560  
QY 1561 GACGGGAGCTGTGGAGGAAGCAGAGGAAGAGCGGCGAGTTTCGAGTTTCGCTCCCAAGAAC 1620  
Db 1561 GACGGGAGCTGTGGAGGAAGCAGAGGAAGAGCGGCGAGTTTCGAGTTTCGCTCCCAAGAAC 1620  
QY 1621 CTGAGGGATTTCAGCCCATTTGTGTTTCAGAGATGACTCTCTCTGAAAGCTGTCTGGGTATCTG 1680  
Db 1621 CTGAGGGATTTCAGCCCATTTGTGTTTCAGAGATGACTCTCTCTGAAAGCTGTCTGGGTATCTG 1680  
QY 1681 AGCCAGGCTTCCAGGCGGCAAGTTTGGACATGAGTGTGAGATCACTGTCTCCCTTGC 1740  
Db 1681 AGCCAGGCTTCCAGGCGGCAAGTTTGGACATGAGTGTGAGATCACTGTCTCCCTTGC 1740  
QY 1741 CATTCGCAACATGAGCATTTCAACCTTGAGACGAGAGTACCTTCCCGATTTCAGGAAC 1800  
Db 1741 CATTCGCAACATGAGCATTTCAACCTTGAGACGAGAGTACCTTCCCGATTTCAGGAAC 1800  
QY 1801 TTACATGAGATGACCTGTGACTTCCATCTCTGAAAGTTGGGTTCGGGTTCAGATTCGGCAC 1860  
Db 1801 TTACATGAGATGACCTGTGACTTCCATCTCTGAAAGTTGGGTTCGGGTTCAGATTCGGCAC 1860  
QY 1861 GCTGTCCCGGATCTCCCGAGAACAGCTTCGCGAGGGTTTCGATCGCCCAACATCAT 1920  
Db 1861 GCTGTCCCGGATCTCCCGAGAACAGCTTCGCGAGGGTTTCGATCGCCCAACATCAT 1920  
QY 1921 CGTCACGCTCGGTTTCATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCACCTCGGGTC 1980  
Db 1921 CGTCACGCTCGGTTTCATCGACCCGCTGTGGCGCATCAAGAGGTTCTTCCACCTCGGGTC 1980  
QY 1981 AGAGGCTCTCTAGCGCAGAGCATCAAGCTCGTGGAGTTCACCTTACAGCTGATCCG 2040  
Db 1981 AGAGGCTCTCTAGCGCAGAGCATCAAGCTCGTGGAGTTCACCTTACAGCTGATCCG 2040  
QY 2041 CCGGAGGAGGCGAGATCGTCGAGGCCCGGCGCAGGCAAAACAGAGAGGATCGTGC 2100  
Db 2041 CCGGAGGAGGCGAGATCGTCGAGGCCCGGCGCAGGCAAAACAGAGAGGATCGTGC 2100  
QY 2101 ACATGACTGTTTCGATTCTTCAGTTTCATCGCTTTCGCGGGATGGACCTGATCTGATG 2160  
Db 2101 ACATGACTGTTTCGATTCTTCAGTTTCATCGCTTTCGCGGGATGGACCTGATCTGATG 2160  
QY 2161 ATTATATATCCGTGTGATCTTGTGAGGCAAAATTAATAATGGGAGATGAAGCAGACATCC 2220  
Db 2161 ATTATATATCCGTGTGATCTTGTGAGGCAAAATTAATAATGGGAGATGAAGCAGACATCC 2220  
QY 2221 TGTACAGGTTTCATCGAGCTAGGCGAGGCCGCGGACGACGCGCGGCTTCGGGAGCAGCA 2280  
Db 2221 TGTACAGGTTTCATCGAGCTAGGCGAGGCCGCGGACGACGCGCGGCTTCGGGAGCAGCA 2280  
QY 2281 AGAGCCTTCGGGAGCTGCTCTCAAATTCCTGATTCGCGGGCGGGAACAGCAGCGCAGCA 2340  
Db 2281 AGAGCCTTCGGGAGCTGCTCTCAAATTCCTGATTCGCGGGCGGGAACAGCAGCGCAGCA 2340  
QY 2341 CGCTGTGCTGTGATCAGCAATGGCCATGTCCCAACCGGAGCTGGCCGAGAGCTGGCC 2400  
Db 2341 CGCTGTGCTGTGATCAGCAATGGCCATGTCCCAACCGGAGCTGGCCGAGAGCTGGCC 2400  
QY 2401 GCGAGCTGTGCGGTTTCGAGGCGGAGCGCGCGGAGGAGGCGCTCGCTGCTGCTG 2460

Db	2401	GCAGCTGTGCGGCTTCGAGCGGAGCGCGCGCGAGAGGGCGTGCCTCGTGCCT	2460	Db	3481	GTCACTCGGATGACGACCAAGCCCAAGCAGTGTAGAGCGAAGCGATCTTTTGGAGTTGG	3540
Qy	2461	GCAGCGGCTGTGACCGCGACGACGAGCGGTTGCGCGCCCGCGTGCAGTTGCGGGCC	2520	Qy	3541	AGACATCTCGGACCCCTTGGGCTCGGAGTCTATCTTCGCTCTCTCTGCTGTGTCG	3600
Db	2461	GCAGCGGCTGTGACCGCGACGACGAGCGGTTGCGCGCCCGCGTGCAGTTGCGGGCC	2520	Db	3541	AGACATCTCGGACCCCTTGGGCTCGGAGTCTATCTTCGCTCTCTCTGCTGTGTCG	3600
Qy	2521	TCCTCACCTACGACAGCGCTCGGCAAGCTGTCTACTTCCAGCTGTGCTCACGAGCG	2580	Qy	3601	TGGCGGACCGCGCCCGCGCTCGTGTTCGACCAATCCCGCGCCCGACCGGTTTCT	3660
Db	2521	TCCTCACCTACGACAGCGCTCGGCAAGCTGTCTACTTCCAGCTGTGCTCACGAGCG	2580	Db	3601	TGGCGGACCGCGCCCGCGCTCGTGTTCGACCAATCCCGCGCCCGACCGGTTTCT	3660
Qy	2581	TCCGCTGTACCCCGCGCTCGGCAAGCTGTCTACTTCCAGCTGTGCTCACGAGCG	2640	Qy	3661	GTACAAACACCTCATCCGCGCGCGCGCGAGTGACACACCGCGGACCGCGTATCAT	3720
Db	2581	TCCGCTGTACCCCGCGCTCGGCAAGCTGTCTACTTCCAGCTGTGCTCACGAGCG	2640	Db	3661	GTACAAACACCTCATCCGCGCGCGCGCGAGTGACACACCGCGGACCGCGTATCAT	3720
Qy	2641	GCACAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG	2700	Qy	3721	CTATAAATCATGTGTGTACTTTTAAACGGCTTTAAACAGGCTTTTATG	3780
Db	2641	GCACAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG	2700	Db	3721	CTATAAATCATGTGTGTACTTTTAAACGGCTTTAAACAGGCTTTTATG	3780
Qy	2701	CCCAAGGGGATCTTGGAGGACGAGTGTCTGCGGACGGGACGAAAGGTGAGGCGGG	2760	Qy	3781	GTAACACAGCTTCAAAATTTGACACAAATTTAAACAGGACAAACCGTAGCTAAACATAAG	3840
Db	2701	CCCAAGGGGATCTTGGAGGACGAGTGTCTGCGGACGGGACGAAAGGTGAGGCGGG	2760	Db	3781	GTAACACAGCTTCAAAATTTGACACAAATTTAAACAGGACAAACCGTAGCTAAACATAAG	3840
Qy	2761	ATGCTGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2820	Qy	3841	AGATGAGAGACAAACCCAAAGTTTAGAGATGAAATAAGCTGAGTAAACAGCAATTC	3897
Db	2761	ATGCTGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2820	Db	3841	AGATGAGAGACAAACCCAAAGTTTAGAGATGAAATAAGCTGAGTAAACAGCAATTC	3897
Qy	2821	CGGAGCTTCCGCGCGGAGCGGTGAGTCAACAGGAGTGGCGGTTCGCAAGCGCTGCGCG	2880	RESULT 3			
Db	2821	CGGAGCTTCCGCGCGGAGCGGTGAGTCAACAGGAGTGGCGGTTCGCAAGCGCTGCGCG	2880	US-10-021-657-1			
Qy	2881	TTCAAGTTTCAAGCGCTTCCAGCGCGGCGGAGATCTGCTGGCAAGAGCTCGCGGTAC	2940	; Sequence 1, Application US/10021657			
Db	2881	TTCAAGTTTCAAGCGCTTCCAGCGCGGCGGAGATCTGCTGGCAAGAGCTCGCGGTAC	2940	; Publication No. US20020083483A1			
Qy	2941	CTGAGATGAGATGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG	3000	; GENERAL INFORMATION:			
Db	2941	CTGAGATGAGATGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG	3000	; APPLICANT: ALBERTSEN, MARC C.			
Qy	3001	GGGACCCGCTGAGTACCGGATGATGATGATGATGATGATGATGATGATGATGATG	3060	; APPLICANT: FOX, TIM			
Db	3001	GGGACCCGCTGAGTACCGGATGATGATGATGATGATGATGATGATGATGATGATG	3060	; APPLICANT: TRIMMELL, MARY			
Qy	3061	CGCGCTCTAGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG	3120	; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND			
Db	3061	CGCGCTCTAGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG	3120	; TITLE OF INVENTION: METHOD OF USING SAME			
Qy	3121	AAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	3180	; FILE REFERENCE: 1148			
Db	3121	AAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	3180	; CURRENT APPLICATION NUMBER: US/10/021,657			
Qy	3181	CCTTAAAAATCTCGTGGGCTTTCGAAACACACATCACTAGTGTGTTTGTACTTACTC	3240	; CURRENT FILING DATE: 2001-12-14			
Db	3181	CCTTAAAAATCTCGTGGGCTTTCGAAACACACATCACTAGTGTGTTTGTACTTACTC	3240	; NUMBER OF SEQ ID NOS: 7			
Qy	3241	CTCAGTGAAGTGTAGTACAGATCAAGTTCATCATATATATATATATATATATATAT	3300	; SOFTWARE: PatentIn Ver. 2.1			
Db	3241	CTCAGTGAAGTGTAGTACAGATCAAGTTCATCATATATATATATATATATATATAT	3300	; SEQ ID NO 1			
Qy	3301	CGATGCTTCCGCGGACCTTTTGGAGACCATTAATGACAGGCGGTGTAAGAAAGGCTTC	3360	; LENGTH: 1906			
Db	3301	CGATGCTTCCGCGGACCTTTTGGAGACCATTAATGACAGGCGGTGTAAGAAAGGCTTC	3360	; TYPE: DNA			
Qy	3361	TTCTGCGGCGAAGTTTGGGTTTCAGAGTCTGTCGCTTTCGAGCAGCAAAAAAGTTTGG	3420	; ORGANISM: Zea mays			
Db	3361	TTCTGCGGCGAAGTTTGGGTTTCAGAGTCTGTCGCTTTCGAGCAGCAAAAAAGTTTGG	3420	; FEATURE:			
Qy	3421	AAGGATCTGAACCTTGAACCGAAATGGCTTCGAAATATGCTCGCATCGGGCGGGGCC	3480	; NAME/KEY: CDS			
Db	3421	AAGGATCTGAACCTTGAACCGAAATGGCTTCGAAATATGCTCGCATCGGGCGGGGCC	3480	; LOCATION: (1..1638, 1642..1767)			
Qy	3481	GTCACTCGGATGACGACCAAGCCCAAGCAGTGTAGAGCGAAGCGATCTTTTGGAGTTGG	3540	US-10-021-657-1			

Query Match 35.4%; Score 1378.4; DB 13; Length 1906;  
Best Local Similarity 82.1%; Pred. No. 0;  
Matches 1833; Conservative 0; Mismatches 16; Indels 385; Gaps 6;

Qy	1088	CATGAGGAGCTCATACACACCGCGGACGCGCATCTTCCCACTAGCAGGGCC	1147
Db	9	CACGAGGAGAGTCTACCTCACCGCGGACGCGCATCTTCCCACTAGCAGGGCC	68
Qy	1148	TCACAGTACATCGGCTCTCTCTCATGATCTTGGTCCAGAGTGGAG	1207
Db	69	TCACAGTACATCGGCTCTCTCTCATGATCTTGGTCCAGAGTGGAG	128
Qy	1208	CCTGAGGAGAGAAAGGCCGAGATCATGGCAGTATCGGTGCAACGGTGGAGCT	1267
Db	129	CCTGAGGAGAGAAAGGCCGAGATCATGGCAGTATCGGTGCAACGGTGGAGCT	188
Qy	1268	GAGGAACTACCCGAGTACGAGTGGCTTCTCGGTACCTGTTCACGCGACAGGACGT	1327
Db	189	GAGGAACTACCCGAGTACGAGTGGCTTCTCGGTACCTGTTCACGCGACAGGACGT	248
Qy	1328	GACCGTCGACATGCCGTTTCACTTCTTACCTACCTACCTGCTGACCCCGGTGAATGTCGAGCA	1387

|||||  
Db 249 GACCGTCGACATGCGGTTCTCTTACACCTACATCGCTGACCGGTTGAATGTCGAGCA 308  
Qy  
1388 TGTCTCAAGACTAACTTCAACAAATTAACCCCAAGGTAAATGACCTGAACCTCACTGATGTT 1447  
Db  
309 TGTCTCAAGACTAACTTCAACAAATTAACCCCAAGGTAAATGACCTGAACCTCACTGATGTT 340  
Qy  
1448 CAGTCTTCGGAAATCAGAGCTGAAGAGTGAATCGAATGTGCTGAACACACCGTGTAGGGAA 1507  
Db  
341 -----AGGGAA 346  
Qy  
1508 TCGGTACAGATCCTACATGAGACGTGCTCTCGGTGAGCGCATCTTCAAGCCGACGCGC 1567  
Db  
347 TCGGTACAGATCCTACATGAGACGTGCTCTCGGTGAGCGCATCTTCAAGCCGACGCGC 406  
Qy  
1568 AGCTGTGAGGAAGCAGAGGAGAGCGCGAGTTTCGAGTTTCGCTTCCAAAGAACCTGAGGG 1627  
Db  
407 AGCTGTGAGGAAGCAGAGGAGAGCGCGAGTTTCGAGTTTCGCTTCCAAAGAACCTGAGGG 466  
Qy  
1628 ATTTGAGCGCCATTGTGTTGAGAGTACTCTGTTGAGAGTACTCTGTTGAGAGTACTGAGCCAGG 1687  
Db  
467 ATTTGAGCGCCATTGTGTTGAGAGTACTCTGTTGAGAGTACTCTGTTGAGAGTACTGAGCCAGG 526  
Qy  
1688 CATCAAGCGCAGGAAAGTGTGAGACATGCAAGGTGAGATCACTGCTCCTTGCATTGCC 1747  
Db  
527 CATCAAGCGCAGGAAAGTGTGAGACATG----- 555  
Qy  
1748 AACATGAGCAATTTCAACCTGAGACACAGAGACTACCTTCCGATTCAGAGAACTTTACATG 1807  
Db  
556 -----CAGGAACCTTTACATG 570  
Qy  
1808 AGGATGAGCTGGACTCCATCTGCAAGGTTGGGTTGCGGTTGAGATCGGCACGCTGTG 1867  
Db  
571 AGGATGAGCTGGACTCCATCTGCAAGGTTGGGTTGCGGTTGAGATCGGCACGCTGTG 630  
Qy  
1868 CCGGATCTCCCGCAGAACAGGTTTCGCGCAGCGGTTTCGATCGCGCAACATCATGTCACG 1927  
Db  
631 CGGATCTCCCGCAGAACAGGTTTCGCGCAGCGGTTTCGATCGCGCAACATCATGTCACG 690  
Qy  
1928 CTGCGTTTCATCGACCCGCTGTGCGCATCAAGAGTTTCCTCCAGCTCGGTTCAAGGCC 1987  
Db  
691 CTGCGTTTCATCGACCCGCTGTGCGCATCAAGAGTTTCCTCCAGCTCGGTTCAAGGCC 750  
Qy  
1988 CTCCTAGCGCAGAGCATCAAGCTCTGAGACGATTCACCTACAGCTGATCGCGCGAGG 2047  
Db  
751 CTCCTAGCGCAGAGCATCAAGCTCTGAGACGATTCACCTACAGCTGATCGCGCGAGG 810  
Qy  
2048 AAGGCCGAGATCGTCTGAGCGCCCGGCGCAGCGCAACAGGAGAGGTACGTGCACATGAC 2107  
Db  
811 AAGGCCGAGATCGTCTGAGCGCCCGGCGCAGCGCAACAGGAGAGGTACGTGCACATGAC 853  
Qy  
2108 TGTTCGATCTTTCAGTTTCATCGTCTTGCGCGGAGTGGACCTGATCTGATTTGATTATAT 2167  
Db  
854 ----- 853  
Qy  
2168 ATCCGTGTGACTTGTGAGGACAAATTAATAATGGGCGAGTGAAGCAGCATCTCTGTCAG 2227  
Db  
854 -----AGATGAAGCAGCATCTCTGTCAG 878  
Qy  
2228 GTTTCATCGAGCTAGCGAGCGCGGAGCAGCAGCGCGCGCTTTCGGGAGCAGCAAGAGCCT 2287  
Db  
879 GTTTCATCGAGCTAGCGAGCGCGGAGCAGCAGCGCGCGCTTTCGGGAGCAGCAAGAGCCT 938  
Qy  
2288 CCGGAGCTGTGTCTCAATCTGATGATCCCGGGCGGGAGCAGCAGCGCGCAGCAGCCTGTC 2347  
Db  
939 CCGGAGCTGTGTCTCAATCTGATGATCCCGGGCGGGAGCAGCAGCGCGCAGCAGCCTGTC 998  
Qy  
2348 GTGGTTCAGCAGATGCGCATGCTCCACCCGAGCGTGGCGGAGAGCTGCGCGCGAGCT 2407  
Db  
999 GTGGTTCAGCAGATGCGCATGCTCCACCCGAGCGTGGCGGAGAGCTGCGCGCGAGCT 1058  
Qy  
2408 GTGGCGGTTTCGAGCGGAGCGCGCGCGGAGGAGCGCTGCGGCTGCTGCCCTGCGGCGG 2467  
|||||

Db 1059 GTCGGCTTCGAGCGGAGCGCGCGCGAGGAGGGGCTCACGCTCGTCTCTGCGCGG 1118  
Qy  
2468 CGTGAAGCGCGAGCAAGCGGTTTCGCGCGCGCGCGCGCGAGTTCGCGGCGCTCTCTCAC 2527  
Db  
1119 CGTGAAGCGCGAGCAAGCGGTTTCGCGCGCGCGCGCGCGAGTTCGCGGCGCTCTCTCAC 1178  
Qy  
2528 CTAGCAGAGCTCGCAAGCTGCTTACCTTCCACGCTGCTTCCACGAGACGCTTCGCGCT 2587  
Db  
1179 CTAGCAGAGCTCGCAAGCTGCTTACCTTCCACGCTGCTTCCACGAGACGCTTCGCGCT 1238  
Qy  
2588 GTACCCCGCGCTCCCTCAGGTGAGCGCGCGCGCAGACGCGACCTCCGCTCCAGAGCAGC 2647  
Db  
1239 GTACCCCGCGCTCCCT----- 1254  
Qy  
2648 ATGCACTGAGTGACCTGAATGCAATGCAATGCACTTGCGCGCGCGCGCAGACCCGAGG 2707  
Db  
1255 -----CAGGACCCGAGG 1267  
Qy  
2708 GGATCTCGAGGAGCAGCTGCTGCGGACGGGACGAGGTGAGGGCGCGCGGATGCTGA 2767  
Db  
1268 GGAATCTCGAGGAGCAGCTGCTGCGGACGGGACGAGGTGAGGGCGCGCGGATGCTGA 1327  
Qy  
2768 CGTACGTCCTCTACTCGATGGGCGGATGAGTACAACTGGGCGCCCGACGCGCGCGAGCT 2827  
Db  
1328 CGTACGTCCTCTACTCGATGGGCGGATGAGTACAACTGGGCGCCCGACGCGCGCGAGCT 1387  
Qy  
2828 TCGGCGCGGAGCGGTGATCAACGAGGATGGCGGTTTCGCGACGCGCTGCGCGCTTCAAGT 2887  
Db  
1388 TCGGCGCGGAGCGGTGATCAACGAGGATGGCGGTTTCGCGACGCGCTGCGCGCTTCAAGT 1447  
Qy  
2888 TCAGGCGGTTTCAGGCGCGGCGGAGGATCTGCTTGGCGCAAGGACTCGGCGCTTCACTGAGA 2947  
Db  
1448 TCAGGCGGTTTCAGGCGCGGCGGAGGATCTGCTTGGCGCAAGGACTCGGCGCTTCACTGAGA 1507  
Qy  
2948 TGAAGTGGCGCTGGCCATCTCTTTCGCTTCTACAGCTTCGCGCTGCTGAGGCGGAGC 3007  
Db  
1508 TGAAGTGGCGCTGGCCATCTCTTTCGCTTCTACAGCTTCGCGCTGCTGAGGCGGAGC 1567  
Qy  
3008 CGGTGAGTACCGCATGATGACCATCTCTTCCATGGCGCGCGCTCAAGTTCGCGCTCT 3067  
Db  
1568 CGGTGAGTACCGCATGATGACCATCTCTTCCATGGCGCGCGCTCAAGTTCGCGCTCT 1627  
Qy  
3068 CTAGGCGCTGCTGATGTCATGGCGATTTG-----GGATATCATCCCGCTTAAATCC----- 3117  
Db  
1628 CTAGGCGCTGCTGATGTCATGGCGATTTGATATGATATGATATCGTCCCGCTTAAATCCAGAC 1687  
Qy  
3118 -----TTAAAAATTTGCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 3163  
Db  
1688 AATAAGCGCTCGTGTACAAATTTGCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1747  
Qy  
3164 TGGTGGCTTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 3223  
Db  
1748 TGGTGGCTTGGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 1807  
Qy  
3224 TGTTCCTTACTTACTCTCTCAGTGGAAAGTGTAGTGACAGCATACAAGTTCATCATATATA 3283  
Db  
1808 TGTTCCTTACTTACTCTCTCAGTGGAAAGTGTAGTGACAGCATACAAGTTCATCATATATA 1857  
Qy  
3284 TTAATCTTCTTCTT 3297  
Db  
1868 TTAATCTTCTTCTT 1881

## RESULT 4

US-10-412-000-1  
; Sequence 1, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND









— E  
3  
C  
a

841 TGGTGGCAGATCACAAAGGAACACAAAGGCTAAGCC

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Db 379 CTGTCNCGGATCTCCCGGAGACAGACTTCNCCAGCGTTTCGATGCGGCTAACATCATC 438  
QY 1922 GTACGCTGCGTTTCATCGACCCGCTGTGGCGATCAAGAGGTTCTTTCCAC 1972  
Db 439 GTACNCTGCGTTTCATCCACCCNCTGTGGCGATCCAGAGTTCTTTCCCC 489

RESULT 8  
US-10-412-000-3  
; Sequence 3, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 494  
; TYPE: DNA  
; ORGANISM: Sorghum sp.  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (1)..(494)  
; OTHER INFORMATION: "n" bases may be a, t, c, g, other or unknown  
US-10-412-000-3

Query Match 6.9%; Score 268.4; DB 14; Length 494;  
Best Local Similarity 77.9%; Pred. No. 3.8e-65;  
Matches 367; Conservative 0; Mismatches 28; Indels 76; Gaps 1;

QY 1502 AGGGAATCGGTGACAGATCTTACATGACGCTGCTCTCGGTGAGCGGATCTTCAAGCCG 1561  
Db 95 AGGGGACGTTGACAGATCTTACATGATGCTCTCTCGGTGAGCGGATCTTCAAGCTG 154  
QY 1562 ACGGCGAGCTGTGGAGGAACAGAGGAAGCGGAGTTTCGAGTTTCGCTTCCAGAAC 1621  
Db 155 ACGGCGAGCTGTGGAGGAACAGAGGAAGCGGAGTTTCGAGTTTCGCTTCCAGAAC 214  
QY 1622 TGAGGGAATTCAGCGCCATTTGTTTCAGAGAGTACTCCCTGAAGCTGTGGGTACTGA 1681  
Db 215 TGAGGGAATTCAGTGCCTCAATGTTTTCAGAGAGTACTCCCTGAAGCTGTGGGTACTGA 274  
QY 1682 GCGAGGATCCAAAGGCGAAGAGTTGTGACATGCGAGGTGAGATCACTGCTCCCTTGC 1741  
Db 275 GTCAGGATCCAAAGGCGAAGAGTTGTGACATG----- 309  
QY 1742 ATTGCCACATGAGCATTTCAACTGAGACAGAGACTACCTTGCAGATTTCAGGAACCT 1801  
Db 310 -----CAGGAACCT 318  
QY 1802 TACATGAGGATGACGCTGGACTCCATCTGCAAGGTTTGGTTTGGGTTCGAGATCGGCA 1861  
Db 319 TACATGAGGATGACACTGGACTGATCTGCAANGTTGGTTTGGGTTCGAGATCGGCA 378  
QY 1862 CTGTCGCGGATCTCCCGGAGAACAGTTTCGCGAGCGGTTTCGATGCGGCAACATCATC 1921  
Db 379 CTGTCNCGGATCTCCCGGAGAACAGCTTCNCCAGCGTTTCGATGCGGCTAACATCATC 438  
QY 1922 GTACGCTGCGTTTCATCGACCCGCTGTGGCGATCAAGAGGTTCTTTCCAC 1972  
Db 439 GTACNCTGCGGTTTCATCCACCCNCTGTGGCGATCCAGAGTTCTTTCCCC 489

RESULT 9

US-10-021-657-6  
; Sequence 6, Application US/10021657  
; Publication No. US20020083483A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/021,657  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 267  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-021-657-6

Query Match 6.7%; Score 263; DB 13; Length 267;  
Best Local Similarity 100.0%; Pred. No. 8.3e-64;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 ATCTATTTTCTTGTGGCAGATCACAAAAAGGAACACAAAGGCTTAAGCCTCTACTTG 889  
Db 5 ATCTATTTTCTTGTGGCAGATCACAAAAAGGAACACAAAGGCTTAAGCCTCTACTTG 64  
QY 890 TTCGGGAGTTAGTGCAGGACACCATATGAATCAAGAAATCTTAATTTGGGTCACACC 949  
Db 65 TTCGGGAGTTAGTGCAGGACACCATATGAATCAAGAAATCTTAATTTGGGTCACACC 124  
QY 950 AAGATTGTCTCTCGAGGTTGGGGGTCCTTAAGGTTGGTAGTACCAATACCCAATATA 1009  
Db 125 AAGATTGTCTCTCGAGGTTGGGGGTCCTTAAGGTTGGTAGTACCAATACCCAATATA 184  
QY 1010 TCACCTAAACACCCCAATCCATGCTACATACATACATACATACATACATACATACAT 1069  
Db 185 TCACCTAAACACCCCAATCCATGCTACATACATACATACATACATACATACATACAT 244  
QY 1070 ACCCTTCATCAAGACACCATCG 1092  
Db 245 ACCCTTCATCAAGACACCATCG 267

RESULT 10  
US-10-412-000-6  
; Sequence 6, Application US/10412000  
; Publication No. US20030182689A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, MARC C.  
; APPLICANT: FOX, TIM  
; APPLICANT: HUFFMAN, GARY  
; APPLICANT: TRIMNELL, MARY  
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES MEDIATING MALE FERTILITY AND  
; FILE REFERENCE: 1148  
; CURRENT APPLICATION NUMBER: US/10/412,000  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/670,153  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 267  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-412-000-6

Query Match 6.7%; Score 263; DB 14; Length 267;  
Best Local Similarity 100.0%; Pred. No. 8.3e-64;  
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 ATCTCATTTTCTTGGTGGCAGATCAAAAGGAAACAAAGGCTTAAGCTTCCTACTTG 889  
|||  
Db 5 ATCTCATTTTCTTGGTGGCAGATCAAAAGGAAACAAAGGCTTAAGCTTCCTACTTG 64  
|||  
QY 890 TTCCGGAGTTAGTTCAGGACACCATATGAATGAAGAAATCTTAATTTGGGCTCACACC 949  
|||  
Db 65 TTCCGGAGTTAGTTCAGGACACCATATGAATGAAGAAATCTTAATTTGGGCTCACACC 124  
|||  
QY 950 AAGATTGCTCTCTCGAGTTGGGGGTCCTTAAGGTTGGTAGTAGCAATACCCAATATA 1009  
|||  
Db 125 AAGATTGCTCTCTCGAGTTGGGGGTCCTTAAGGTTGGTAGTAGCAATACCCAATATA 184  
|||  
QY 1010 TCACCTTAACAAACCCCAATCCATGCTACATACATACATACATACATACATACATG 1069  
|||  
Db 185 TCACCTTAACAAACCCCAATCCATGCTACATACATACATACATACATACATACATG 244  
|||  
QY 1070 ACCCTTCATCAAGAGACCATGG 1092  
|||  
Db 245 ACCCTTCATCAAGAGACCATGG 267  
|||

## RESULT 11

US-10-425-114-23429  
; Sequence 23429, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 23429

; LENGTH: 1807  
; TYPE: DNA

; ORGANISM: Zea mays  
; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3596-027-C6\_FLI  
US-10-425-114-23429

Query Match 4.2%; Score 164.2; DB 12; Length 1807;

Best Local Similarity 67.6%; Pred. No. 4.2e-35;

Matches 248; Conservative 0; Mismatches 113; Indels 6; Gaps 1;

QY 2693 CCGAGACCCCAAGGGATCTGGAGGACGAGTCTGCGGACGGGACGAGGTGAGGG 2752

Db 1213 CTCGTGATAACAGCAGTCTTCGGAGACGAGTCTGCTGCTGACGGTTCAGCGTCAGCA 1272

QY 2753 CCGGGGATGTGAGTACGTACGTCCTACTCGATGGGCGGATGGATCAACTGGGGC 2812

Db 1273 AGGAGACATCGTGTCTACGTCCCTACGCCATGGCGGATGAGTACCTCTGGGGCA 1332

QY 2813 CCGAGCGGCGAGCTTCCGGCGGAGCGGTGGATCAACAGGATGGCGGTTCGGCAAG 2872

Db 1333 AGGACCGGAGGTGTTCGGCGGAGCGGTGGATGATGAGGAGGTTCAGACAG 1392

QY 2873 CGTCCCGTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTTCAAGTT 2932

Db 1393 AGAGCCCGTTCAAAATTCAGAGGTTCCAGGCTGGTCCGAGGATATGCTTGGGAGGAGT 1452

QY 2933 CCGCGTACCTGAGATGAGATGGCGCTGCCATCTCTTGGCTTCTACAGCTTCGGC 2992

Db 1453 TCGCGTACCGGAGATGAGATGCTTCGCGCGGCTGCTCTCCGATTTCTTCGTTCGGC 1512

QY 2993 TCGTGGAGGGCAC-----CCGGTGCAGTACCGGATGATGACCATCTCTCCATGGCG 3046

|||

Db 1513 TCCGCGACGGCGACAAAGGAGCGTGAATTACCGACCATGATCAGCTCCACATCGACG 1572

QY 3047 ACGGCT 3053

Db 1573 AGGGTCT 1579

## RESULT 12

US-10-425-114-35451

; Sequence 35451, Application US/10425114  
; Publication No. US2004003488A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 35451

; LENGTH: 825  
; TYPE: DNA

; ORGANISM: Zea mays  
; FEATURE:

; OTHER INFORMATION: Clone ID: UC-ZMROB73033D09\_FLI  
US-10-425-114-35451

Query Match 4.0%; Score 154.2; DB 12; Length 825;

Best Local Similarity 65.1%; Pred. No. 1.6e-32;

Matches 244; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

QY 2698 GACCCCAAGGGATCTCGAGACGACGCTGCTGCGGACGGGACGAGGTGAGGCGCGC 2757

Db 210 GACCAACAAGGAGTCTGGAGACGAGGTTCCTCCGACGGGACGGTCTGAAGAAGGCG 269

QY 2758 GGGATGTTGACGTACGTGCTCTACTCGATGGGCGGATGGAGTCAACTGGGGCCCCGAC 2817

Db 270 ACCAAGTGTACTACGCCATGTACTTCCATGGGCGGATGGAGGATCTGGGGCCAGCAGC 329

QY 2818 GCGGCGAGTCTCCGGCGGAGCGGTGGATCAACAGAGATGGCGGCTTCGCAACGCGTCG 2877

Db 330 TGCCCGAGTACAAGCCGAGCGGTGGCT---CCGGGACGGACGCTTCGTGGCGAGTCC 386

QY 2878 CGGTTCAAGTTCAAGCGGCTTCAGCGCGGCGGAGGATCGCTGGGCAAGGACTCGCGG 2937

Db 387 GCCTCAAGTTCAAGCGGCTTCAGCGCGGCGGCGGCTTCGCTCGGCAAGGACTTCGCGC 446

QY 2938 TACCTGCAGATGAAGATGGCGCTGGCCATCTCTTGCCTTCTACAGCTTCGCGCTGCTG 2997

Db 447 TACTACAGATGAAGTTCAAGCGGCTTCAGCGCGCTCCATCTCCGCGCTACCGGTTCGCTC 506

QY 2998 GAGGGGACCGCGTGCAGTACCGCATGATGACCATCTCTCCATGGCGCAGCGCTCAAG 3057

Db 507 GAGGGGACCGCGTGCAGTACCGCATGATGACCATCTCCATGATGAGCGGCTCAAG 566

QY 3058 GTCCGCTCTAGG 3072

Db 567 GTGACGCTCACCAAG 581

## RESULT 13

US-10-238-1236

; Sequence 1236, Application US/10260238  
; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.  
; APPLICANT: Moughamer, Todd G.  
; APPLICANT: Briggs, Steven P.

APPLICANT: Cooper, Bret  
APPLICANT: Glazebrook, Jane  
APPLICANT: Goff, Stephen A.  
APPLICANT: Katagiri, Fumiaki  
APPLICANT: Kreps, Joel  
APPLICANT: Provart, Nicholas  
APPLICANT: Ricke, Darrell  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
FILE REFERENCE: 60111-NP  
CURRENT APPLICATION NUMBER: US/10/260,238  
CURRENT FILING DATE: 2002-09-26  
PRIOR APPLICATION NUMBER: US 60/325,448  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: US 60/325,277  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: US 60/370,620  
PRIOR FILING DATE: 2002-04-04  
NUMBER OF SEQ ID NOS: 6077  
SEQ ID NO 1236  
LENGTH: 1617  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: N region  
LOCATION: (1326)..(1326)  
OTHER INFORMATION: n = any nucleotide  
US-10-260-238-1236

Query Match 3.9%; Score 151.6; DB 15; Length 1617;  
Best Local Similarity 62.7%; Pred. No. 1.5e-31;  
Matches 235; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 2697 GCACCCCAAGGGATCTCGAGGACGACGTCTCCGCGACGGGACGAGGTGATGGCGG 2756  
Db 1218 GGAGCAACAAGCGCGGTCCGCGACGACGTCTCCGCGACGGGACGAGGTGATGGCGG 1277

Qy 2757 CGGGATGTGACGTACGTGCCCTACTCGATGGGCGGATGAGTACACTGGGCCCCGA 2816  
Db 1278 CGACAGGTGTGTGTCTACTACTCCATGGGAGGATGAAGCGGGTNTGGGCAAGGA 1337

Qy 2817 CGCGCGAGCTTCGCGCGGAGCGGTGGATCAAGAGGATGGCGGTTCGCAACGCGTC 2876  
Db 1338 CTGACGGAGGTTCAGCGCGGAGCGGTGGATCAAGAGGATGGCGGTTCGCAACGCGTC 1397

Qy 2877 GCGGTTCAAGTTCAGCGGTTCCAGGCGGCGGCGGAGGATCTGCTCGGCAAGGACTCGGC 2936  
Db 1398 GTCGAACAAGTTCGCGGTTCAAGTTCGCGGCGGCGGAGGATCTGCTCGGCAAGGACTCGGC 1457

Qy 2937 GTACTCGAGATGAAGATGGCGGTGGCCATCTCTTGGCTTCTACAGCTTCGCGCTGCT 2996  
Db 1458 GCTGTCGAGATGAAGTTCAGCGCGGCGGCGGATGGCTGGGACTTTCGCGGTGGAGTGGT 1517

Qy 2997 GGAGGGGACCGGTGACGTACCGCATGATGACCATCTCTCCATGGGCGACGCGCTCAA 3056  
Db 1518 GCCGGGCGACGTCTCGGAGCGAGGCTTCGCTATCTACTCCATGAGATGGGCTCTT 1577

Qy 3057 GGTCCGGCTCTCTAG 3071  
Db 1578 GGTAAAGGTCAAGAG 1592

## RESULT 14

US-10-260-238-4619  
Sequence 4619, Application US/10260238  
Publication No. US20040016025A1  
GENERAL INFORMATION:  
APPLICANT: Budworth, Paul R.  
APPLICANT: Moughamer, Todd G.  
APPLICANT: Briggs, Steven P.  
APPLICANT: Cooper, Bret  
APPLICANT: Glazebrook, Jane  
APPLICANT: Goff, Stephen A.

## RESULT 15

US-10-425-114-28465  
Sequence 28465, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 28465

APPLICANT: Katagiri, Fumiaki  
APPLICANT: Kreps, Joel  
APPLICANT: Provart, Nicholas  
APPLICANT: Ricke, Darrell  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
FILE REFERENCE: 60111-NP  
CURRENT APPLICATION NUMBER: US/10/260,238  
CURRENT FILING DATE: 2002-09-26  
PRIOR APPLICATION NUMBER: US 60/325,448  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: US 60/325,277  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: US 60/370,620  
PRIOR FILING DATE: 2002-04-04  
NUMBER OF SEQ ID NOS: 6077  
SEQ ID NO 4619  
LENGTH: 854  
TYPE: DNA  
ORGANISM: Triticum aestivum  
US-10-260-238-4619

Query Match 3.8%; Score 146.4; DB 15; Length 854;  
Best Local Similarity 61.6%; Pred. No. 2.8e-30;  
Matches 234; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 2693 CGCAGGACCCCAAGGGGATCTCGAGGACGACGTCTGCGGACGGGACGAGGTGAGGG 2752  
Db 56 CGATGGAGCAACAAGGGCGTGGTGGCGGAGCGCTGCCGAGCGGCGACGAGGTGGCGC 115

Qy 2753 CCGCGGGATGGTACGTACGTCCCTACTCGATGGGCGGATGGAGTCAACTGGGGCC 2812  
Db 116 CGGGGACAAAGATCATGTGTCTGTACGCGATGGGAGGATGGAGCGGTGTGGGGCA 175

Qy 2813 CCGAGCGGCGAGCTTCGCGCGGAGCGGTGGATCAACAGGATGGCGCTTCGCAACG 2872  
Db 176 AGACTGCGGGAGTTCCGCGCGGAGCGGTGGATCGGGAGGACGGCAACCGCGGTACG 235

Qy 2873 CGTCCGCGTTCAAGTTACGCGCTTCAGGCGGGCGGAGGATCTGCTCGGCAAGGACT 2932  
Db 236 TGCGGTGTAACAAGTTCGTGTCTTCAACTCCGCGCGGAGCATGCTTCGCAAGGACA 295

Qy 2933 CGCGGTACCTGACAGTGAAGATGGCGTGGCCATCTCTTTCGCGCTTACAGCTCCGGC 2992  
Db 296 TGGCGTTCTGTCAGTCAAGCGGTGGCGCGCCCGCGTGGTGAAGAACTTCGAGGTGAGG 355

Qy 2993 TGTGGAGGGGACCCCGGTGCAGTACCGCATGATGACCATCTCTCCATGGCGCGCGCC 3052  
Db 356 CCGTGGCGGACAGCTGCTGGAGCCCCCAAGATCTCCATCATCTCTCCATGAAGAACGGCT 415

Qy 3053 TCAAGGTCCGCTCTTAGG 3072  
Db 416 TCAAGGCCAGGATCAAGAG 435

```
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4757-030-F2_FLI
US-10-423-114-28465

Query Match      3.7%; Score 143.8; DB 12; Length 1947;
Best Local Similarity 52.1%; Pred. No. 2.8e-29;
Matches 402; Conservative 0; Mismatches 352; Indels 17; Gaps 3;

QY 1764 CTGAGACACAGAGACTACCTTCCCGATTACAGAACTTTACATGAGGATGACGCTGGACT 1823
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 525 CCGCGCGCGCGCGGAGAGCTAGACATGACAGAGCTGCTGATGCTTTCGACGCTGGACT 584
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1824 CCATCTGCAAGGTTGGGTTCCGGGTCGAGATCGGACGCTGTCGCGGATCTCCCGGAGA 1883
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 585 CCATCTTCAAGGTCGGGTTCCGGGTCAGCTGGGCGTGTCTGCCGCTCCACGAGGAAG 644
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1884 ACAGC---TTCCGCGAGCGGTTTCGATGCCGCCAACATCATCGTACGCTGGGTTTCATCG 1940
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 645 GCGCGCATTCGCCAAGGCTTCGACGAGCCAGCGAGGTCGTGCGCGCTTCTTGG 704
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1941 ACCCGCTGTGGCGCATCAAGAGGTTCTTCCAGTCGCGGTCAGAGCCCTCTTAGCGCAGA 2000
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 705 ACCCGTTCTGGAAGGCCAAGAGGTTCTGAACCTTCTCGTCCGAGCGGCCATGAAGCGCT 764
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2001 GCATCAAGCTCGTGGACGAGTTCACCTACAGGCTGATCCGCCGAGGAGGCGCGAGATCG 2060
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 765 CGTCCGACCATCAACGGCTTCGTCTACGCGCTCTCGACAGGAAGA----- 812
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2061 TCGAGGCCCGGCGCAGCGGCAACAGAGAGAGTACGTGCACATGACTGTTTCGATTCTT 2120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 813 TCGAGCAGATGGAAGAGACCCACAGGATTTGTAGCCCGCAGAGATATTACCTTCCGT 872
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2121 CAGTTATCTGCTTGGCGCGGATGGAAGCTGATTCCTGATTTGATTTATATCCGTGTGACTT 2180
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 873 AGTCGTACATGATGGGTGCAGTGCAGACGGGACGACGGGATCACGAGGAGATTACAG 932
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2181 GTGAGGACAAATTAATGGGCGGATGAGCAGACATCTCTGTACGGTTTATCGAGCTA 2240
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 933 TTCTGACACGGTGACATTTCCCTGACACGCGCAAGAAAGAGGACATCTCTGTGAGGTTCT 992
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2241 GCGAGGCCGCGGACGACGCGCGGCTTCCGGGACGACAAAGAGCTCCGCGGACGTGCTG 2300
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 993 GGTGAGCGGAGAG--GGATCCCGGCTGCTCCGACACAAAGTACTCCGCGGACATCATC 1050
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2301 CTCAACTTCGTGATCGCGCGGAGACACGACGCGGACGACGCTGTCTGTGTTTACGCGAC 2360
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1051 CTCAACTTCGTGATCGCGCGGCGGACACGACCGCGGGGACCTGTCTGTGTTTCTCTAC 1110
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2361 ATGCCCATGTCCACCGGAGCTGCGGAGAGCTGCGCGGAGCTGTGCGGTTGAG 2420
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1111 GTGCTCTGAGGAACAGAGATCCAGACAGGATTCGCGAGGAGGTGCGCGCGCCACC 1170
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2421 GCGAGCGCGCGCGGAGGCGGTTCGCGCTCGTTCGCTCGCGCGGCGCTGACGCGGAC 2480
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1171 TCGGCGGACGCGCGCTCGCGCTCCAGAGCTGTCTGCGGTGCTGACCGAGAGCGCCATC 1230
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2481 GACAAAGGTTTCGCGCGCGCGGTTCGCGGAGTTTCGCGGCGCTTCCTACCTAC 2531
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1231 AGCAAGATGCACTACTGTCACGCGCGGTGACGAGAGACCTTCGCGCTCTAC 1281
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: March 7, 2004, 19:35:33  
Job time : 910 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2004, 22:04:59 ; Search time 150 Seconds  
(without alignments)  
7051.576 Million cell updates/sec

Title: US-10-021-657-1

Perfect score: 1906  
Sequence: 1 gaattcgacagagggaagc.....aaaaaaaaaaaaaacctcgag 1906

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271.6	14.2	1626	3	US-09-158-767-14
2	271.6	14.2	2181	3	US-09-158-767-10
3	73	3.8	1926	4	US-09-249-585A-4
4	73	3.8	1931	2	US-09-130-114-2
5	70.6	3.7	1365	4	US-09-252-991A-289
6	70.6	3.7	1941	4	US-09-252-991A-248
7	70.6	3.7	2076	4	US-09-252-991A-238
8	69.2	3.6	1551	4	US-09-252-991A-5309
9	69.2	3.6	4158	4	US-09-252-991A-5348
10	69.2	3.6	4933	4	US-09-252-991A-5227
11	68.6	3.6	1866	4	US-09-615-192A-103
12	68.6	3.6	1866	4	US-09-169-789-103
13	68.6	3.6	2013	4	US-09-615-192A-404
14	67.6	3.5	1563	4	US-09-252-991A-15934
15	67.6	3.5	1933	4	US-09-252-991A-15763
16	67.6	3.5	3351	4	US-09-252-991A-15871
17	64.2	3.4	1212	4	US-09-252-991A-14382
18	64.2	3.4	1512	4	US-09-252-991A-14321
19	63.8	3.3	1566	6	5290690-3
20	63.8	3.3	77536	4	US-09-410-551B-1
21	63.6	3.3	15872	3	US-09-105-537-1
22	63.6	3.3	15872	4	US-09-091-609-1
23	63.6	3.3	15872	4	US-09-091-609-3
24	63.4	3.3	20235	1	US-07-642-734C-3
25	63.4	3.3	20235	3	US-08-439-009A-3
26	62.8	3.3	2049	4	US-09-252-991A-505
27	62.6	3.3	77536	4	US-09-410-551B-1

c	28	62.4	3.3	453	4	US-09-252-991A-275	Sequence 275, Appl
	29	62.4	3.3	1128	4	US-09-622-439-1	Sequence 1, Appli
	30	62.2	3.3	1182	1	US-07-640-476-4	Sequence 4, Appli
	31	62.2	3.3	1185	6	5290690-1	Patent No. 5290690
	32	62.2	3.3	2938	3	US-09-072-917A-8	Sequence 8, Appli
	33	62	3.3	1269	4	US-09-252-991A-8698	Sequence 8698, Ap
	34	62	3.3	1944	4	US-09-252-991A-9081	Sequence 9081, Ap
c	35	61.4	3.2	1248	3	US-09-105-537-7	Sequence 7, Appli
	36	61.4	3.2	13613	3	US-09-105-537-3	Sequence 3, Appli
c	37	61.2	3.2	1485	1	US-08-474-601-23	Sequence 23, Appl
	38	61.2	3.2	1485	1	US-08-474-556-23	Sequence 23, Appl
	39	61.2	3.2	1485	1	US-08-479-382-23	Sequence 23, Appl
	40	61.2	3.2	1485	1	US-08-470-354-23	Sequence 23, Appl
	41	61.2	3.2	1485	1	US-08-479-383-23	Sequence 23, Appl
	42	61.2	3.2	1485	2	US-08-479-041-23	Sequence 23, Appl
	43	61.2	3.2	1485	3	US-08-819-646-23	Sequence 23, Appl
	44	61.2	3.2	1485	4	US-09-195-716-23	Sequence 23, Appl
	45	60.4	3.2	1401	2	US-08-812-412-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1

US-09-158-767-14

; Sequence 14, Application US/09158767A

; Patent No. 6180363

; GENERAL INFORMATION:

; APPLICANT: Batard, Yannick

; APPLICANT: Durst, Francis

; APPLICANT: Schalk, Michel

; APPLICANT: Warck-Reichhart, Daniele

; TITLE OF INVENTION: RECODING OF DNA SEQUENCES PERMITTING

; TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST

; FILE REFERENCE: A32000

; CURRENT APPLICATION NUMBER: US/09/158,767A

; CURRENT FILING DATE: 1998-09-23

; EARLIER APPLICATION NUMBER: FR 97-12094

; EARLIER FILING DATE: 1997-09-24

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 14

; LENGTH: 1626

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Altered sequences

US-09-158-767-14

Query Match	14.2%;	Score 271.6;	DB 3;	Length 1626;
Best Local Similarity	54.5%;	Pred. No. 4.3e-47;	Indels 90;	Gaps 6;
Matches	736;	Conservative 0;	Mismatches 524;	
QY	276	CACCTACATCGCTGACCGGTGAATGTGAGCATGTCTCAAGACTPACTTACCAATTA	335	
Db	243	CACCGTCACCTGCGACCGCGCAACCTGAGCAGCTCTGAGGCGCGCTTCGACAACTA	302	
QY	336	CCCAAGGAATGCTGTACAGATCTTACATGAGACGTCTCTCGGTGACGGCATCTTCA	395	
Db	303	CCCAAGGCGCCCTTCTGCGACGGCGCTCTTCCGGGACCTGCTCGGCGACGGCATCTTCA	362	
QY	396	CGCGGCGCGGAGCTGTGGAGGAGCAGAGGAGCGGAGTTTCGAGTTGCGCTTC---	452	
Db	363	TTCCGACGGCGACACCTGCTGCGCAGCGCAAGCGCGCGCTCAGTTTACCAACCG	422	
QY	453	CAAGAACTCTGAGGATTTTCAGCGCCATTGTGTTTACAGAGTACTTCCCTGAAGCTGCGG	512	
Db	423	CACGCTCCGACGGCCATGTCCGCTGGGTCTCGCGCTCCATCCACGGCGCTCTCTGCC	482	
QY	513	TATACCTGAGCCAGGCTTCAAGTTCGAGCAGCGAAAGTTGTGGACATCGCAGAACTTTACATGAG	572	
Db	483	CATCTCTGGCGACGCGCGCAAGCGGAGCGGAGTGGATCTCCAGGACCTCTCTCTCCG	542	

QY	573	GATGACGCTGGACTCCATCTCTCCAGAGTTTGGGTTCCGGGTCCAGATCCGACGCTGTGCGC	633
Db	543	CCTCACCTTGGACAACATCTCTCGGCCCTGGCCTTCCGCAAGACCCGGAGACGCTCGGCCA	602
QY	633	AGATCTCCCGAGAACAGCTTCGGCGCAGGCGTTCCATGCCGCCCAACATCATCATCACGCT	692
Db	603	GGGCTTCGCGAGAACGAGTTTCGCTCCGGTTTCGACCGGCCACCGAGGCCACGCTCAA	662
QY	693	CGGTTTCATCGAACCG-----CTGTGGCGCATCAAGAGGTTTTCACGTCGGGTCCAGA	746
Db	663	CCGCTTTCATCTTCCCGGAGTTCTGTGTGGCGTGC AAAAAGTGGCTGGGCTTCGCGCATGGA	722
QY	747	GGCCCTCTCTAGCGCAGAGCATCAAGCTCTGTGGCAGAGTTCACTTACAGCGTGATCCCGCG	806
Db	723	GACCACGCTGACACGACGATGGCCCACTGCAACAGTACTCTCGCCCGTTCATCAAGAA	782
QY	807	GAGGAAGCCGAGATCGTCCAGGTCCTGGGCCAGCGGCAAAACAGAGAGAAGATGAACACGA	866
Db	783	GCGCAAGCTCGAGCTCGCCGCCGCAACGGCAATGCGACACGCGCGCAGCAGCAGCA	842
QY	867	CATCCTGTCAAGGTTTATCGAGCTGGGCGAGGCCGCGACGACGCGCGCGCTTCGGGGA	926
Db	843	CCTGCTCTCCCGGTTTCATCGGAAGGTTTCCTATCTCGGACG-----	883
QY	927	CGATPAGAGCCTCCGGGACGTGTGTGCTCAACTTCGTGATCCCGCGGGGACACGACGCG	986
Db	884	-----AGTCGCTCCAGCAGCTGGCGCTCAACTTCATCCTCGCGCGCGCGACACCTCCTC	938
QY	987	GACGACGCTGTCTGTGTTTACGACATAGGCCATGTCTCCACCGGACGTGGCGGAGAAGCT	1046
Db	939	CGTGGCGCTCTCTGTGTTTCTTGCTGTGTCTCACCCACCTCGCTGGAGCGCAAGAT	998
QY	1047	GCGCGCGGAGCTGTGCGGTTTCGAGGCGGAGCGCGCGCGGAGGAGGCGTTCACGCTCGT	1106
Db	999	CGTGGCGAGCTCTGCTC-----	1019
QY	1107	GCTCTGGCGGCGGCTGACGCCGACGACAAAGCGTTTCGCGCCGCGTGGCGAGTTTCGC	1166
Db	1020	TCTCGCCGCGTACGCGGGCGCCATGACCCGGCAT-----GTGCGCTGGC	1064
QY	1167	GGGCTCCTCACCTACGACGCTCGGCAAGCTGGTCTACTCTCACGCTTCGCTCACGGA	1226
Db	1065	GGAGCCCTTCACCTTCGAGAGCTCGACCGGCTGGTCTACTCTCAAGCGCGGCTGTCCGA	1124
QY	1227	GAGCTCCGCTGTACCCCGCGCTCCCTCAGGACCCCAAGGGATCCTGGAGGACGACGT	1286
Db	1125	GACCTCCGCTCTACCCCTCCGTCGCCGAGGACTTCAAGCACGTCGTCGCGGACGACTA	1184
QY	1287	GCTGCGCGAGCGGACGAAAGTGTAGGCGCGGGGAGTGGTAGCTACGTGCCCTACTCGAT	1346
Db	1185	CTTCCCGCAGCGGCACCTTCGTGCGCGCGGGTCTCGGTCACTTACTCCATATCTCGGC	1244
QY	1347	GGGCGGATGGAGTACAACCTGGGGCCCGCAGCGCGGAGCTTCGCGCGGAGCGGTGCGAT	1406
Db	1245	GGGCGCATGAGAGGGGTGTGGGGGAGGACTGCTCTGAGTTCCGGCCCGAGCGATGGCT	1304
QY	1407	CAACGAGATGGCG-----TTTCCGCAACGGCTGCGCTTCAAGTTACGCGTTTCCAGGC	1463
Db	1305	GTGCGCGAGCGGACCAAGTTTCGAGCAGCAGCACTCTGTAAGTTCTGGCGTTTCAACGC	1364
QY	1464	GGGCGCGAGATCTGCCTGGCGCAGGACTCGGGGTACTCTCAGATGAGATGGCGCTGGC	1523
Db	1365	CGGCGCAGGGGTGTGCTTGGCGAAGGACCTTAGCCCTACTCTCAGATGAGAAACATCGCCGG	1424
QY	1524	CATCTCTTCGCTTCTACAGCTTCCCGGCTGCTGGAGGGGCACCGGTCGAGTACCGCAT	1583
Db	1425	GAGCGTCTGCTCGGCACCGCTGACCGTGGCGCGGCCACCGCTGGAGCAGAGAT	1484
QY	1584	GATGACCATCTCTTCATGGCGGACGGCT	1613
Db	1485	GTGCTCACGCTCTTCTATGAGGCGGGCT	1514

```

RESULT 2
US-09-158-767-10
; Sequence 10, Application US/09158767A
; Patent No. 6180363
; GENERAL INFORMATION:
; APPLICANT: Batard, Yannick
; APPLICANT: Durst, Francis
; APPLICANT: Schalk, Michel
; APPLICANT: Werck-Reichhart, Daniele
; TITLE OF INVENTION: RECORDING OF DNA SEQUENCES PERMITTING
; TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST
; FILE REFERENCE: A32000
; CURRENT APPLICATION NUMBER: US/09/158,767A
; CURRENT FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: FR 97-12094
; EARLIER FILING DATE: 1997-09-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: fastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 2181
; TYPE: DNA
; ORGANISM: Trifolium aestivum
US-09-158-767-10

```

Query Match	14.2%; Score 271.6; DB 3; Length 2181;
Best Local Similarity	54.5%; Pred. 4.6e-47;
Matches	736; Conservative 0; Mismatches 524; Indels 90; Gaps 6
Qy	276 CACCTATACGCTCACC CGGTGAATGCGACATGTCTCAGACTAACTTCACCAATTA 335
Db	354 CACCGTCACTGCGACCCCGGCAACCTTGGAGCAGCTCTGAAGGCGCGCTTCGACAATTA 413
Qy	336 CCCCAAGGGGAATCGTGTACAGATCCTATCATGGACGTGCTCCCTGGTGACGGCATCTTCAA 395
Db	414 CCCCAAGGGCCCTTCTGGCACGGGCTCTCCGGGACCTGTCTGGGACGGAATCTTCAA 473
Qy	396 CGCCGACGGGAGCTGTGGAGGAAGACAGAGGAACGCGGAGTTTCGAGTTTCGGCTC --- 452
Db	474 TTCCGACGGGACACCTTGGCTCGCGCAGCGCAGACGGCCCGGCTTCGAGTTTCACACCGG 533
Qy	453 CAAGAACTTGAGGATTTTCAGCGCCATTTGTTTTCAGAGAGTACTCCCTGTAAGCTGTCCGG 512
Db	534 CACGCTCCGACGGCCCATGTCCCGCTGGGTCTCCGCTTCCATCCACGGCCGCTCTCTGCC 593
Qy	513 TATACTGAGCCAGGCATCCAAAGGCAGGCAAGTTGTGGACATGCAGGAACCTTTATCATGAG 572
Db	594 CATCTTGGCCGACGCGGCCAAGGCGCAGGTGGATCTCCAGGACCTCTCTCTCCG 653
Qy	573 GATGACGCTGGACATCCATCTTGCAAGTTGGGTTTCGGGTTCCAGATTCGACAGCTGTCCGC 632
Db	654 CCTCACCTTCGACAACATCTGCGGCTTGGCTTCGGCAAGGACCGGAGACGCTCGCCCA 713
Qy	633 AGATCTCCCGGAGAACAGCTTCGCGCAGGCGTTCCGATCCGCGCAACATCATCACGCT 692
Db	714 GGGCTCTCCGGAGAACAGTTGGCTTCGGGTTTCGACCGGCCACCGAGGCCAGCTCAA 773
Qy	693 GCGGTTTCATCGACCCG-----CTGTGGCGCATCAAGAGTTCTTCACAGTCGGGTGAGA 746
Db	774 CCGCTTCATCTTCCCGAGTTCTGTGTGGCTGCAAAAGTTGGCTGGGCTTCGGCATGGA 833
Qy	747 GGGCCCTCCTAGCGCAGAGCATCAAGCTCGTGGACGAGTTTCACTTACAGCGTGTATCCGCG 806
Db	834 GACCACGCTGACCAGCAGCATGGCCACGTCGACCACTA CCTCGCCCGCTCATCAAGAA 893
Qy	807 GAGGAAGCCGAGATCGTTCGAGTTCGGGCGCAGGGCAACAGAGAAAGATGAAGCAGGA 866
Db	894 GCGCAAGCTCGAGCTCGCCCGCGCAACGCGCAATTCGCACACGCGCGCAGCACGACGA 953
Qy	867 CATCTCTCACGTTTCATCGAGCTGGCGAGGCGCGGACGACGCGCGGCTTCGGGGA 926
Db	954 CTTGCTCTCCGGTTTCATCGGGAAGGTTTCTTACTCGGACG----- 994
Qy	927 CGATAAGAGAGCTCCGGGACGTGTGTCTCAACTTCGTATCGCCGGGCGGGACAGCGGC 986



Db 995 -----AGTCGCTCCAGACGTGGCGCTCACTTCATCTCCGCGCGCGACACCTCC 1049  
QY 987 GACGACGCTCTGTCGTTACGACACATGGCCATGTCCACCCGAGAGTGTCGCCGAGAGCT 1046  
Db 1050 CGTGGGCTCTCTGTCGTTCTTCTGGCTCGTGTCCACCCACCTGCGTGAGCGCAAGAT 1109  
QY 1047 GCGCCGCGAGCTGTGCGCTGTTCGAGGCGGAGCGCGCGCGGAGGCGCTCACGCTCGT 1106  
Db 1110 CGTGGCGAGCTGTCTC-----CGT 1130  
QY 1107 GCTCTGCGCGCGGCTGACCGCGACGACAAAGGGTTTCGCGCGCGTGGCGAGTTCCG 1166  
Db 1131 TCTCGCGGCTACGCGGCGCCCATGACCCGCGANT-----GTGGCTGGC 1175  
QY 1167 GCGCCTCTCTACCTACGACAGCTCGGACAGCTGTGTCTACCTCAGCGCTGCGTCAACGA 1226  
Db 1176 GGAGCCTCTTACCTTCAAGAGCTCGACCGCTGGTCTAATCAAGCGCGCGCTGTCCGA 1235  
QY 1227 GAGCCTCGGCTCTACCCCGCGCTCCCTCAGGACCCCAAGGGATCTTGAGGACAGCT 1286  
Db 1236 GACCTTCGCGCTTACCTTCCCTCCGCGCGCGGTGCTCGTCACTTCTGCTCGGACGACTA 1295  
QY 1287 GCTGCGCGGACGGAAGTGAGGCGCGCGCGGATGGTGACGTACGTGTCCTTACTCGAT 1346  
Db 1296 CTTCCCGCGCGGACCTTCTGTCGCGCGCGGTGCTCGTCACTTCTGCTCGGCTCGGCT 1355  
QY 1347 GGGCGGATGGAGTCAACTGGGCGCGCGCGCGGAGGTTCCGCGCGGAGCGGTGGAT 1406  
Db 1356 GGGCGGATGAAGGGGCTGTGGGCGGAGGACTGCTCGAGTTCCGCGCGGCGGCTGGCT 1415  
QY 1407 CAACGAGGATGGGCG---GTTCCGCAACGCGTCCGCGTTCAAGTTCACGCGCTCCAGGC 1463  
Db 1416 GTGCGCGGACGACCAAGTTTCAGCAGCAGCAGTCTGTACAGTTCTGTCGCTTCAACGC 1475  
QY 1464 GGGCGCGAGGATCTGCTGGGCAAGACTCGGCGTACCTGCAATGAAGATGGCGTGGC 1523  
Db 1476 CGGCGCGAGGCTGCTGCGCAAGGACCTAGCTTACCTGCAATGAAGATCGCCGC 1535  
QY 1524 CATCTCTTCGCTTCTACAGTTCGCGTGTGAGAGGCGGACCCGCTGAGTACCGAT 1583  
Db 1536 GAGCGTGTCTCCGCAACGCTGACCGTGGCGCGCGGCGCACCGGCTGGAGCAGAAGAT 1595  
QY 1584 GATGACCATCTCTCCATCGCGCACGCGCT 1613  
Db 1596 GTGCTCAGCTCTTCAAGAGGCGGGCT 1625

## RESULT 3

US-09-249-585A-4/c

; Sequence 4, Application US/09249585A

; Patent No. 6417002

; GENERAL INFORMATION:

; APPLICANT: Horlick, Robert

; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES

; FILE REFERENCE: 0867/0905

; CURRENT APPLICATION NUMBER: US/09/249,585A

; CURRENT FILING DATE: 1999-02-11

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: Patent version 3.0

; SEQ ID NO 4

; LENGTH: 1926

; TYPE: DNA

; ORGANISM: Epstein Barr Virus

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)..(1926)

; OTHER INFORMATION: template strand of EBNA-1 DNA

US-09-249-585A-4

Query Match

Best Local Similarity 3.8%; Score 73; DB 4; Length 1926;

Matches 345; Conservative 0; Mismatches 410; Indels 6; Gaps 2;

QY 737 TGGGTACAGAGCCCTCTTAGCGCAGAGCATCAAGCTCTGTGAGCAGATTCACCTACAGCG 796  
Db 1014 TGGGGCCGAGGTGACGAGGAGCTTGGGGCCGAGGTGGAGGACGAGGACGCGGAGGACG 955  
QY 797 TATCCGCGCGAGGAGGCGGAGATCGTGTGAGTCCGCGCCAGCGCAACACGAGAAGA 856  
Db 954 AGGACGGGAGGAGGACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGACG 895  
QY 857 TGAAGCACAACATCTCTGTACCGTTTATCGAGTGGCGGAGGCCGCGCAGCAGCGCGGG 916  
Db 894 AGGACGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGACGCGGAGGACGAG 835  
QY 917 GCTTCGGGACGATTAAGACCTTCCGGGACGTGGTGTCTCACTTCGTGATCGCGGGCG 976  
Db 834 ACGGGAGGACCGGGAGGAGGACGAGGACGCGGAGGAC-----GGGAGGACGAGGAC 780  
QY 977 ACAACGACGCGGACGCGCTGTGTGTTCACGACATGCGCATGTCCCAACCGCGAGCTGG 1036  
Db 779 GGAGGACGGGAGGACGAGGACGCGGAGGACGCGGAGGACGAGGAC-GGGAGGAGCAG 721  
QY 1037 CCGAGAGCTGCGCCGCGAGCTGTGCGGTTTCGAGCGGAGCGCGCGCGGAGGAGGCG 1096  
Db 720 ACGGGAGGACGAGGACGCGGAGGACGAGGACGCGGAGGACGCGGAGGACGCGGAGG 661  
QY 1097 TCAAGCTCTGTCTCTGCGCGCGCTGACCGCGACGACGAGGCTTCCGCGCGCGCTGG 1156  
Db 660 ACGAGGACGGGAGGACGAGGACGCGGAGGACGCGGAGGACGCGGAGGACGAGGAC 601  
QY 1157 CGCAGTTCCGCGCGCTCTCCTACCTACGACAGCTCGGCAAGCTGTCTACCTCCAGCT 1216  
Db 600 AGGAGGACGAGGACGCGGAGGACGCGGAGGACGCGGAGGACGAGGACGCGGAGGAC 541  
QY 1217 GCGTCAACGAGAGCTTCCGCTGTATCCCGCGCTTCCCTCAGGACCCCAAGGGGATCT 1276  
Db 540 GCGAGGAGGACGAGGACGCGGAGGAGGACGAGGACGCGGAGGACGCGGAGGACGCG 481  
QY 1277 AGGACGAGCTGTCTGCGGACGCGGACGAGGTCGAGGTCGCGCGCGGATGTGTAGCT 1336  
Db 480 AGGACGAGGACGCGGAGGACGCGGAGGAGGACGAGGACGCGGAGGAGGACGAGGAC 421  
QY 1337 CCTACTCGATGGCGCGGATGAGTACAACTGGGCGCCCGACGCGCGAGCTTCCGCGCG 1396  
Db 420 AGGACGCGGAGGACGCGGAGGAGGACGAGGACGCGGAGGACGAGGACGCGGAGGAC 361  
QY 1397 AGCGGTGAGTCAACGAGGATGCGGCTTCCGCAACGCGTTCGCGCTTCAAGTTTCA 1456  
Db 360 GGGAGGACGCGGAGGACGCGGAGGAGGACGAGGACGCGGAGGAGGACGAGGACGCG 301  
QY 1457 TCCAGCGCGCGCGGAGGATCTGCTTGGGCAAGGACTCGGCG 1497  
Db 300 ACGAGGACGGGAGGCGGAGGACGAGGACGAGGACGAGGACGAGGACGAGGACGAG 260

## RESULT 4

US-09-130-114-2/c

; Sequence 2, Application US/09130114

; Patent No. 5976807

; GENERAL INFORMATION:

; APPLICANT: Horlick, Robert A.

; APPLICANT: Dama, Bassam B.

; APPLICANT: Robbins, Alan K.

; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes

; TITLE OF INVENTION: From Multiple Transfected Episomes

; FILE REFERENCE: 0867/ID903US1

; CURRENT APPLICATION NUMBER: US/09/130,114

; CURRENT FILING DATE: 1998-08-06

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq For Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 1931

; TYPE: DNA

; ORGANISM: EBNA

US-09-130-114-2

Query Match 3.8%; Score 73; DB 2; Length 1931;  
Best Local Similarity 45.3%; Pred. No. 3.3e-06;  
Matches 345; Conservative 0; Mismatches 410; Indels 6; Gaps 2;

QY 737 TCGGGTTCAGAGCCCTCCTAGCGCAGAGCATCAAGCTCGTGACGAGTTTCACCTACAGCG 796  
DB 1014 TGGGCGGAGGTGACGGAGGAGCTGGGCCCGAGGTGAGGACGAGGACGGGAGGACG 955  
QY 797 TGATCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 856  
DB 954 AGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 895  
QY 857 TGAAGCAGACATCTCTCAGCTTCATCGCTGAGCTGGGAGGAGGAGGAGGAGGAGG 916  
DB 894 AGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 835  
QY 917 GCTTCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 976  
DB 834 ACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
QY 977 ACACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1036  
DB 779 GAGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 721  
QY 1037 CCGAGAGCTGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1096  
DB 720 ACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 661  
QY 1097 TCACGCTCGTCTCGCGGCGGCTGACCGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 1156  
DB 660 ACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 601  
QY 1157 CGCATTCGCGGCGCTCTCTCACTACGAGAGCTCGGCAAGCTGCTACCTCCAGCT 1216  
DB 600 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 541  
QY 1217 GCGTACCGAGAGCTCGCGCTGTACCGCGGCTCCCTCAGGACCCGAGGAGGAGG 1276  
DB 540 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 481  
QY 1277 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1336  
DB 480 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 421  
QY 1337 CCTACTCGATGGGCGGATGAGATCAACTGGGGGCGGAGCGCGGCGGAGCTTCCG 1396  
DB 420 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 361  
QY 1397 AGCGTGGATCAACAGGATGCGGCTTCGCGCAACGCTGCGCTCAAGTTTCAGGCGT 1456  
DB 360 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301  
QY 1457 TCCAGCGGGGCGGAGGATCTGCTGGGCAAGGAGCTCGGCG 1497  
DB 300 ACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 260

## RESULT 5

US-09-252-991A-289/c  
; Sequence 289, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 289

; LENGTH: 1365

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-289

Query Match 3.7%; Score 70.6; DB 4; Length 1365;  
Best Local Similarity 46.5%; Pred. No. 9.6e-06;  
Matches 229; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 753 CCTAGCGCAGAGCATCAAGCTCGTGAGCGAGTTTCACCTACAGCGTGATCCGCGCGAGGAA 812  
DB 985 CCACGCCGAAACGATACAGGCGCTCGTCAATTCACCGCATCTGGCGTCCGCTGCTGGA 926  
QY 813 GCGGAGATCTCGAGGTCCGCGGCGAGCGGCAACAGAGAGATGAAGCAGCAGCATCCT 872  
DB 925 GGGGAGAGCCGTCGACTACGCGCGCAAGCACATCCAGGTGAAGGCGCCAACTGCTCTA 866  
QY 873 GTCAGGTTTCATCGAGCTGGCGGAGGCGGCGAGCAGACGCGCGCGGCTTCGGGGAGATAA 932  
DB 865 TCCGCGCTTCAGCAGCGCGCTCCGCGCTGCTACTTCGCGCGCTCCTCGAGGCGCGCCA 806  
QY 933 GAGCTCCGCGACGTGTGCTCAACTTCGTGATCGCGGCGGCGAGCAGCAGCGAGCGAC 992  
DB 805 GGACTCCGCGCGAGCAGGTTCGAGCTGTACTGACCTGGGCGGAGCGCGCGCGCGCGGT 746  
QY 993 GCTGCTGTGTTTCAGCAGCATGGGCGATGTCACCGCGAGCAGTGGCGGAGAGCTGCGCG 1052  
DB 745 TGCCGAGAAGATGCCCGAGGTACCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCTT 686  
QY 1053 CGAGCTGTCGCTTCGAGCGGAGCGCGCGCGGAGGCGCGCGCGCGCGCGCGCGCTG 1112  
DB 685 CGGATCCGCTGACGCTGATCGTCCGCGAGACCGAGGAGGAGGAGGAGGAGGAGGAG 626  
QY 1113 CGGCGCGCTGACCGCGAGCAAGGCGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1172  
DB 625 TCGACTGATCGCGCACTTCGAGCAGACACCATCGCGCGCGCGCGCGCGCGCGCGCG 566  
QY 1173 CCTCAGCTACGAGCGCTTCGCAAGCTGCTTACCTCCAGCGCTGCGTCAACGAGAGCT 1232  
DB 565 CTTGCACTCGTTCGCGCAGCAACGATGCGCGCGCTGCAAGCGGCTTCGCGCGAGCACT 506  
QY 1233 CCGCTCTACCCC 1245  
DB 505 GGAAGTCAGCCCC 493

## RESULT 6

US-09-252-991A-248

; Sequence 248, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 248

; LENGTH: 1941

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-248

Query Match

Best Local Similarity 46.5%; Pred. No. 1e-05;

3.7%; Score 70.6; DB 4; Length 1941;



Db	564	TGAACGTCGGCAACCTCGAAACAGCTGCACCTGATCACCCGAGCGCCCAAGCTCAACGCCA	623
Qy	989	CGACGCTGTCGTGGTTCCAGCACATGCCCATGTCACCACCGGAGCT--GGCCGAGAGC	1045
Db	624	AGCTCTACGCGAAGAACTCAACATCTCACGGCCGCAACGAGCTCCAGGCCGACAGCC	683
Qy	1046	TGGCCGCGAGCTGTGCGCGTTTGAGCGGAGCGCGCGCGAGGAGGGGCTCACGCTCG	1105
Db	684	TGCAGGCCACGCGCGCGCGCGGATGGCAGCGAGAAAGCCACAGCTGGCGATCGACAGCT	743
Qy	1106	TGCTCTCGCGGGCGGCTGAAGCGGACAGAGCGGCTTGCGCGCGCGCGTGGCGGAGTTCG	1165
Db	744	CGCGCTTGGCGGGGATACCGCGGGCGGATCGCGCTGGTCGGCACCGAGCGGGCGCTGG	803
Qy	1166	CGGGCTTCCTCACTAGACAGCCTCGGCAAGCTGGTCTACCTCCAGCGCTGCGTCAACG	1225
Db	804	GGTTCGCGCTGGCCGGCGCATGGCCGCCACAGCGCGCGCATCCGCATCGACGCCACGG	863
Qy	1226	AGACGCTCCGCTGTATCCCGCGCGTCCCTCAGGACCCCAAGGGGATCCTCGAGGAGCAGC	1285
Db	864	GCAAGCTAGGCTGGCCCGCGGCTCCAGCCAGGCGGACCTGAAGATCGCGGCCAGGCCG	923
Qy	1286	TGCTGCGGACGGGACGA	1303
Db	924	TGGAGCTGAACGGCAGA	941

```

RESULT 9
US-09-252-991A-5348
; Sequence 5348, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5348
; LENGTH: 4158
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5348

```

Query Match	3.6%;	Score 69.2;	DB 4;	Length 4158;
Best Local Similarity	45.6%;	Pred. No. 2.4e-05;		
Matches	282;	Conservative 0;	Mismatches 333;	Indels 3; Gaps 1;
QY	689	CGCTGCGGTTTCATCGACCCGCTCTGGCGCATCAAGAGGTTCTTCCACGTGGGTCTCAGAGG	748	
Db	404	CGCAGGTGATCTCTCAACCAAGGTCAACCGCGGGCAACCGCACACCTCTGGCGGCTACACCG	463	
QY	749	CCCTCTTAGCGCAGAGCATCAAGCTCTGTGGACAGTTTCACCTACAGGGTGATCCGCCGGA	808	
Db	464	AGGTGCCCGGGCAGTCTGGCGCGGGTGATCTGTGCCAACCCGCAACCGCATCACCTGCCAGG	523	
QY	809	GGAAGCCGAGATCTGTCGAGGTCTGGGCGCCAGCGGCAACAGAGAGAGATGAACACAGCA	868	
Db	524	GCTGCGGCTTCATCAACACGCGCGCGGACCTCTACCACCGCAGCCGATCATCGACG	583	
QY	869	TCTGTCTACGGTTTCATCGAGCTGGGGCAGGCCCGGCAACGACGGCGCGGCTTCGGGGACG	928	
Db	584	GCCAGACGCTTGGAGCGCTTCCAGGTGGAACGGCGCGGACATCGTCTCGAAGGCGCCGAAC	643	
QY	929	ATAAGAGCCTCCGGGACGTGGTCTCAACTTCGTGTATCGCCGGCGGGACACAGACGGCGGA	988	
Db	644	TGAACTTCGGCAACCTTCGAAACAGTTTCGACTGATACCCGCGAGCGCCAGGCTCAACGCCCA	703	

QY	989	CGAGCGTGTGTGGTTTCA	CGACATGCGCCATGTCC	CACCCGGACGT---	GGCCGGAAGC	1045
Db	704	AGCTCTACGCGAAGAAC	CTCAACATGTCACCGG	CCGCAAGCGTCCAG	CGCGACAGCC	763
QY	1046	TGCGCCCGGAGCTGTGT	GCGCGTTTGAGCGGAG	CGCGCGCGGAGGGCG	TCACGCTCG	1105
Db	764	TGCAGGCGACGCGCG	CGCGCGGATGGCAG	CGAGAAGCCACAGCT	GGCGATCGACAGCT	823
QY	1106	TGCTCTCGGCGGGCGT	GACCGCAGACAAAGG	CGTTCGCGCCGCGT	TGGCGCGAGTTCG	1165
Db	824	CGGCGCTGGGCGGAG	TGTACCGCGGGCGAT	CGCCCTGTGTCGCA	CCGACGAGCGGGCGTGG	883
QY	1166	CGGCGCTTCCTCAC	TACGACAGCCTCGG	CAAGCTGTGTCTAC	CTTCAGCGCTCGGTCA	1225
Db	884	GGGTGCGGCTGGCG	GGGCAATGGCGCCAG	CGCGGGCGGACAT	CCGCATCGACGCCAGCG	943
QY	1226	AGAAGCTCCCGCTGT	ATCCCGCGGTCCCT	CAGGACCCCAAGGGG	ATCTTGGAGGACGACG	1285
Db	944	GCAAGCTGACCTGG	CGCCAGGCGCTC	CAGCCAGGCGGAC	CTTGAGATCGCGGCCAGGCGG	1003
QY	1286	TGCTGCCGGA	CGGGAAGA			1303
Db	1004	TGGAGCTTGAAC	CGGCAAGA			1021

```

RESULT 10
US-09-252-991A-5227/c
; Sequence 5227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.1136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5227
; LENGTH: 4953
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5227

```

Query Match	3.6%	Score 69.2	DB 4	Length 4953
Best Local Similarity	45.6%	Prod. No. 2.5e-05		
Matches 282	Conservative 0	Mismatches 333	Indels 3	Gaps 1

  

QY	689	CGCTGGCGTTCA	TCGACCGCCTGTGGCGCATCAAGAGGTTCTTCCAGTTCGGGTCA	GAGG	748
Db	4595	CGCAGGTGATCTTCA	ACCAGGTCACCGCGGCGCAACCGCAGCACCTTGGCGGCTACACCG		4536
QY	749	CCCTCTTAGCGCAGAGCATCA	AAGCTCGTGGACAGATTACCTACAGGTCGATCGCGCGGA	808	
Db	4535	AGGTGGCGGGCATTC	CGCGCGGGTGATCGTCCCAACCGCGCAGGCATCACCTGGCCAGG	4476	
QY	809	GGAAGGCGGAGATCTG	CTCAGGTCGGGCGCAGCGGCAACAGAGGAGATGAAGCAGCAGCA	868	
Db	4475	GCTGCGGCTTCATCA	CACGCGCGCGGCAACCTTCAACCACGGCAAGCCGATCATGACCG	4416	
QY	869	TCCTGTTCAGGTTTCAT	CTGAGCTGGGCGCAGCGCGGCACGACGGCGCGCGCTTCGGGGACG	928	
Db	4415	GCCAGCGCTTGGAGCGCT	TCCAGTGGACGGCGGCGACATCTGTCGTAAGGCCCGCGAAC	4356	
QY	929	ATAAGACCTTCGGGAC	GTGGTGTCTCAACTTCGTGATCGCGGCGGGGACACACGACGGCGGA	988	
Db	4355	TGAACGTGGACACCTCG	AACAGTTCGACTGATCACCCGACGCGCCAGCTCAACGCCCA	4296	
QY	989	CGACGCTGTGCTGGTTCA	CGGCATGCGGCATGTCCACCCCGGACGTT---GGCCGGAAGC	1045	

Db 4295 AGCTTACGCGAGAACCTCAACATCGTACCGCGCCGACAGAGTCCAGCGCCGACAGCC 4236  
QY 1046 TGGCCCGGAGCTGTGGCGGTTGAGCGGAGAGCGCGCGGAGGAGGCGGTACGCTCG 1105  
Db 4235 TGCAGGCGACCGCGCGCGCGCGATGGCAGCGAGAACCCACAGCTGGCGATCGACAGCT 4176  
QY 1106 TGCTCTCGCGCGCGCTGAGCGCGCGACACAGGCGTTTCGCGCGCGCGGTGGCGAGTTGG 1165  
Db 4175 CGGCGCTGGGCGGATGATGACCGCGGCGATCCGCTGGTCGGACACGACAGGCGGTGG 4116  
QY 1166 CGGCGCTCTCACTACGACAGCGCTCGGCAAGCTGTGTCTACCTCCAGCGCTGGGTACCG 1225  
Db 4115 GGGTGGCGCTGGCGCGCGACATGGCGCCAGCGCGCGCGACATCCGATCGACGCGCAGCG 4056  
QY 1226 AGAGCTCCCGCTGTACCGCGCGCTCCCTCAGACACCGCAAGGAGATCTCGAGAGACGAG 1285  
Db 4055 GCAAGCTGAGCGCTGGCGCGCGCTCCAGCGCGAGCGCTGAAGATGGCGCGCGCGCG 3996  
QY 1286 TGCTGGCGGACGGGACGA 1303  
Db 3995 TGGAGCTGAACGCAAGA 3978

## RESULT 11

US-09-615-192A-103  
; Sequence 103, Application US/09615192A  
; Patent No. 6410718  
; GENERAL INFORMATION:  
; APPLICANT: Bloksberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; FILE OF INVENTION: Modification of Plant Lignin Content  
; FILE REFERENCE: 11000.1003c4U  
; CURRENT APPLICATION NUMBER: US/09/615,192A  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; PRIOR APPLICATION NUMBER: US 09/169,789  
; PRIOR FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 405  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 103  
; LENGTH: 1866  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-615-192A-103

Query Match 3.6%; Score 68.6; DB 4; Length 1866;  
Best Local Similarity 47.0%; Pred. No. 2.6e-05;  
Matches 352; Conservative 0; Mismatches 384; Indels 13; Gaps 4;  
QY 894 CGAGCGCGGACGAGCGCGCGGCTTCGGGGACGATAGAGCCTCCGGGAGCTGTGCT 953  
Db 705 CAAGACGAGTTCGGTGGAGGGGATCAAGATGTGATCCGACATGGTGGACATCTGCT 764  
QY 954 CAACCTCGTATCGCGCGGAGACGAGCGGACGAGCGCTGTGTGTTCAGCACAT 1013  
Db 765 GGCCTTCTACGCGAGCAAGAGGTGAACGAGTCCGACGATTTGACAACTCGATCAG 824  
QY 1014 GGCCATGTCCCAACCGCGAGTGGCGGAGAGTGGCGCGGAGCTGTGGCGCTTCGAGGC 1073  
Db 825 GCTACGAGAGACAACTCAAGGCCATCATCTGGACGTATGTTCCGGCGGAGCGAGAC 884  
QY 1074 GAGCGCGCGCGAGGAGGCGGTTCACGCTGTGTCTTCGGCGGCGCTGAGCGCGACGA 1133  
Db 885 TGTGGCGTGGCTATCGAGTGGCCATGGCGGAGCTCATGCGAGCCCGGAGCACTGNA 944  
QY 1134 CAAGGGCTTCGCGCGCGCGGTGGCGGAGTTTCGGGCGCTTCCTCACTACGA-----CAG 1187  
Db 945 GAAGGTCCAGCAAGAACTCGCGGATGTCTGTGGCGCTTAGACCGGAGAGTTCGAGAGCGGA 1004

QY 1188 CCTCGCAAGCTGGTCTACTCTCACGCGCTGGTTCACCGAGACGCTCGCTGTGTACCCCGC 1247  
Db 1005 CTTTCGAGAGCTGACCTATCTCAAGTGTCTCAAGAGACACCTCCGCTCCACCCGCGC 1064  
QY 1248 CGTCCCTCAGGACCCCAAGGGGATCTCGAGGAGCAGCTGTCTCGGAGCGGAGCGAGGT 1307  
Db 1065 GATCCCGCTGCTCTCCACGAGACGCGCAGAGACGCCGTGATCTC---CGGCTACCGCAT 1121  
QY 1308 GAGGGCGCGGGGATGGTACGCTACGTGCTTACTCGATGGGCGGATGAGTACCAACTG 1367  
Db 1122 CCGCGCACGCTCCCGGTCATGATCAATGATCAATGGGCCATCGGGCG---TGACCCCGGCTC 1178  
QY 1368 GGGCGCGGAGCGGCGAGCTTCGGCGCGAGCGGTGGATCAAGAGAGTGGCGGTTCGG 1427  
Db 1179 GTGGACCGAACCTGACAAAGTTCAAACCCGCTCCCGGTTCTTGGAGTCAGGATGCCGACTA 1238  
QY 1428 CAACGCTCGCGGTTCAAGTTTCAAGCTTCAAGGCTTCCAGGGCGGCGAGGATCTGCTGGGCA 1487  
Db 1239 CAAGGGGACAACTTCGAGTTTATCCCTTCGGTGGCGCGGAGTCTGTGCCAGGAT 1298  
QY 1488 GGAATCGGCTACTGTCAGATGAAGATGGCGTGGCCATCTCTTCCGCTTCTACAGCTT 1547  
Db 1299 GCAGCTCGGCTCTACGCGCTCGACATGGCGCTGGGCCACCTCTGCACTGCTTACGTTG 1358  
QY 1548 CCGGCTGTGGAGGGGCGCCCGTGCAGTACCGCATGATGACCATCTCTCATGGCGCA 1607  
Db 1359 GGAATCGCGCGGAGTGAAGCGGAGATGAGCATGGGCGAGCTTTC---GGGCTCA 1417  
QY 1608 CGGCTCTAAGTTCGCGCTCTCTAGGCGCG 1636  
Db 1418 CCGCGCGGAGGTCACCGCGCTCGTGGCG 1446

## RESULT 12

US-09-169-789-103  
; Sequence 103, Application US/09169789  
; Patent No. 6653528  
; GENERAL INFORMATION:  
; APPLICANT: Bloksberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; FILE OF INVENTION: Modification of Plant Lignin Content  
; FILE REFERENCE: 11000.1003c2  
; CURRENT APPLICATION NUMBER: US/09/169,789  
; CURRENT FILING DATE: 1998-10-09  
; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; NUMBER OF SEQ ID NOS: 185  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 103  
; LENGTH: 1866  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-169-789-103

Query Match 3.6%; Score 68.6; DB 4; Length 1866;  
Best Local Similarity 47.0%; Pred. No. 2.6e-05;  
Matches 352; Conservative 0; Mismatches 384; Indels 13; Gaps 4;  
QY 894 CGAGCGCGGACGAGCGCGCGGCTTCGGGGACGATAGAGCCTCCGGGAGCTGTGCT 953  
Db 705 CAAGACGAGTTCGGTGGAGGGGATCAAGATGTGATCCGACATGGTTCGACGATCTGCT 764  
QY 954 CAACCTCGTATCGCGCGGAGACGAGCGGACGAGCGCTGTGTGTTCAGCACAT 1013  
Db 765 GGCCTTCTACGCGAGCAAGAGGTGAACGAGTTCGAGACTTCGATCAG 824  
QY 1014 GGCCATGTCCCAACCGCGAGTGGCGGAGAGTGGCGCGGAGCTGTGGCGCTTCGAGGC 1073  
Db 825 GCTACGAGAGACAACTCAAGGCCATCATCTGGACGTATGTTCCGGCGGAGCGAGAC 884

QY	1074	GGAGGGCGCGCGAGGAGGCGTCA	CGCTCTCTGCGGCGCGCTGACGCCGACGA	1133
Db	885	TGTGGCGTTCGGCTATCGAGTGGGCAT	TGGCGAGCTCATGCGAGCCCCGAGGACCTGAA	944
QY	1134	CAAGGCGTTCCGCCGCCGCTGGCGCAGTT	TGCGGGCTCTCTACCTACGA-----CAG	1187
Db	945	GAAGGTCAGCAAGAACTCGCGGATGT	GTGGGGCTTAGACGGGAGATCTCAGGAGAGCGA	1004
QY	1188	CCTGGCAAGTGTCTACCTCCACGCTGGGT	CACCGAGAGCTCCGCTGTACCCCGC	1247
Db	1005	CTTCGAGAGCTGACCTATCTCAAGTGT	GTGCTCAAGAGACCCTCGCTTCACCCGCC	1064
QY	1248	CGTCCCTCAGGACCCCAAGGGATCT	TGGAGGACGAGCTGTGCCGACGGGACGAAGGT	1307
Db	1065	GATCCCGTGTCTCTCCACGAGACGGCAG	AGAGACCGCGTGATCTC---GGCTACCGCAT	1121
QY	1308	GAGGGCGGGGATGTGAGTACGTGCCCT	TACTCGATGGGCGGATGGAGTACAACCTG	1367
Db	1122	CCCCGACGGTCCGGGTCTGATCAAT	TGATGGGCCATCGGGG---TGACCCCGGCTC	1178
QY	1368	GGGCCCGGACGGCGCGAGCTTCCGGC	CGGAGCGGTGGATCAACGAGATGGCGCTTCG	1427
Db	1179	GTGACCGAAGCTGACAAAGTTCAAC	ACGTCCTCGGTTCTGGAGTCAGGATGCCGACTA	1238
QY	1428	CAACGCTCGCGTTCAAGTTCAAGCGTT	CCAGGGCGCCGAGGATCTGCTGGGCAA	1487
Db	1239	CAAGGGAGCAACTTCGAGTTTATCCCT	TTTCGGGTCCGGCGGAGTCTGTGCCAGGAT	1298
QY	1488	GGACTCGGCTACCTGCAGATGAAGAT	TGGCGCTGGCCATCTCTTCCGCTTCTACAGCTT	1547
Db	1299	GCAGCTCGGCTCTAACCGCTTCGACAT	TGGCGGTGGCCACCTCTCTGACTGCTTACGTT	1358
QY	1548	CCGGCTCTGGAGGGGACCCGGTGCAGTA	CCCGCATGATACCATCTCTCCATGCGCA	1607
Db	1359	GGAATCCCGCAGGGATGAAGCCGACG	AGATGGACATGGGCGACGTCTTC-GGGCTCA	1417
QY	1608	CGGCTCAAGGTCGGGCTCTCTAAGGCCG		1636
Db	1418	CCGGCGGAGGTCCACCCGGCTCGTGGCG		1446

RESULT 13

```

US-09-615-192A-404
; Sequence 404, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615.192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 404
; LENGTH: 2013
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-615-192A-404

Query Match      3.6%;      Score 68.6;  DB 4;      Length 2013;
Best Local Similarity 47.0%;      Pred. No. 2.7e-05;
Matches 352; Conservative 0; Mismatches 354; Indels 13; Gaps 4

924 CGAGCGCGGACGACGCGCGCGCGCTTCGGGGACGATPAGAGCTCCGGGACGTGCTGCT 953

```



Query Match 3.5%; Score 67.6; DB 4; Length 1563;  
Best Local Similarity 45.5%; Pred. No. 4.1e-05;  
Matches 281; Conservative 0; Mismatches 334; Indels 1; Gaps 1;  
QY 689 CGCTGCGGTTTATCGACCCCGCTGTGGCGCATCAAGAGTTTCTTCCAGCTGGGTCAGAGG 748  
DB 324 CGCAGGTGATCTTCAACACAGGTACCCGCGGCAACCCGACACCTTGGCGGCTACACCG 383  
QY 749 CCCTCTAGCGCAGAGATCAAGTCTGTGAGAGATTCACCTACAGCGTGTATCCGCGGA 808  
DB 384 AGTGGCCGGGAGTCCGCGGGTGTATCTGCGCAACCCGACCGGATCACCTGCCAGG 443  
QY 809 GGAAGGCCGAGATCGTCCAGTCCGCGGCAACCAAGAGAGATGAAGCAGACA 868  
DB 444 GCTCGGCTTCAACACACCGCGCGGACCTTCAACACCGGCAACCGGATCATGAGCG 503  
QY 869 TCCTGTCAAGTTCATCAAGTGTGGCGAGCGCGCGGAGAGAGAGTTCGCGGAGCG 928  
DB 504 GCCAGCGCTTGGAGCGCTTCCAGGTGACGCGCGGAGATCTGCTCGAAGGCGCGAAC 563  
QY 929 ATAAGAGCCCTCCGGAGCTGTGCTCAACTTCGTGATCGCGGCGGAGACAGCGCGGA 988  
DB 564 TGAAGTCCGCAACCTCGAAGTTCAGCTGATCAACCGCGCGCAAGCTCAACGCCA 623  
QY 989 CGAGCTGTCTGTGTTTCAAGCAGATGGCCATGTCCACCCCGGAGCT---GGCGGAGAGC 1045  
DB 624 AGCTTACGCGAAGAACCTCAACATGTACCGGCGCAACGAGTCCAGGCGGACAGCC 683  
QY 1046 TGGCGCGAGCTGTGGCGCTTTCAGCGGAGCGCGCGGAGAGGCGGTCAACGCTCG 1105  
DB 684 TGCAGGCCACGCGCGCGCGCGATGGCAGCGAGAGACACAGTGGCGATCGACAGCT 743  
QY 1106 TGCTCTGCGCGCGCTGAGCGCGAGACAGAGCGTTCGCGCGCGCGGTGGCGAGTTGG 1165  
DB 744 CGCGCTGGCGGGAGTGTACCGCGGGCGATCCGCTGTGCGCACCGAAGCGGCGTGG 803  
QY 1166 CGGCGCTCTCACTACGACAGCTCGGCAAGCTGTCTTACCTCCAGCGCTGGGTCAACG 1225  
DB 804 GGGTGGCGTGGCGCGGACATGGCGCGGAGCGCGGCGGAGATCCGATCGACCGCAGCG 863  
QY 1226 AGAGCTTCGCTGTACCGCGCTTCCTCAGACCCCAAGGGGATCTTGGAGGAGCG 1285  
DB 864 GCAAGTGTAGCTTGGCGCGCTCCAGCGCGGAGCTTCAAGATCGCGGCCCGCGCG 923  
QY 1286 TGCTGCGGAGCGGAGCA 1303  
DB 924 TGGAGTGAACGCAAGA 941

RESULT 15  
US-09-252-991A-15763/c  
; Sequence 15763, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIORITY FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15763  
; LENGTH: 1953  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15763

Query Match 3.5%; Score 67.6; DB 4; Length 1953;  
Best Local Similarity 45.5%; Pred. No. 4.3e-05;

Matches 281; Conservative 0; Mismatches 334; Indels 1; Gaps 1;  
QY 689 CGCTGCGGTTTATCGACCCCGCTGTGGCGCATCAAGAGTTTCTTCCAGCTGGGTCAGAGG 748  
DB 1595 CGCAGGTGATCTTCAACACAGGTACCCGCGGCAACCCGACACCTTGGCGGCTACACCG 1536  
QY 749 CCCTCTAGCGCAGAGATCAAGTCTGTGAGAGATTCACCTACAGCGTGTATCCGCGGA 808  
DB 1535 AGTGGCCGGGAGTCCGCGGGTGTATCTGCGCAACCCGACCGGATCACCTGCCAGG 1476  
QY 809 GGAAGGCCGAGATCGTCCAGTCCGCGGCAACCAAGAGAGATGAAGCAGACA 868  
DB 1475 GCTCGGCTTCAACACACCGCGCGGAGCTTCAACACCGGCAACCGGATCATGAGCG 1416  
QY 869 TCCTGTCAAGTTCATCAAGTGTGGCGAGCGCGCGGAGAGAGTTCGCGGAGCG 928  
DB 1415 GCCAGCGCTTGGAGCGCTTCCAGGTGAGCGGCGGAGATCTGCTCGAAGGCGCGAAC 1356  
QY 929 ATAAGAGCCCTCCGGAGCTGTGCTCAACTTCGTGATCGCGGCGGAGACAGCGCGGA 988  
DB 1355 TGAAGTCCGCAACCTCGAAGTTCAGCTGATCAACCGCGCGCAAGCTCAACGCCA 1296  
QY 989 CGAGCTGTCTGTGTTTCAAGCAGATGGCCATGTCCACCCCGGAGCT---GGCGGAGAGC 1045  
DB 1295 AGCTTACGCGAAGAACCTCAACATGTACCGCGCGCAACCGAGTCCAGGCGGACAGCC 1236  
QY 1046 TGGCGCGGAGCTGTGCGGCTTTCAGGCGGAGCGCGCGCGGAGGCGGTCAACGCTCG 1105  
DB 1235 TGCAGGCCACGCGCGCGCGCGATGGCAGCGAGAGACACAGTGGCGATCGACAGCT 1176  
QY 1106 TGCTCTGCGCGCGCTGAGCGCGAGACAGAGCGTTCGCGCGCGCGGTGGCGAGTTGG 1165  
DB 1175 CGCGCTGGCGGGAGTGTACCGCGGGCGATCCGCTGTGCGCACCGAAGCGGCGTGG 1116  
QY 1166 CGGCGCTCTCACTACGACAGCTTCGCGCAAGCTGTCTTACCTCCAGCGCTGGGTCAACG 1225  
DB 1115 GGGTGGCGTGGCGCGGAGATGGCGCGGAGCGGCGGAGATCCGATCGACCGCAGCG 1056  
QY 1226 AGAGCTTCGCTGTACCCCGCGCTTCCTCAGGACCCCAAGGGGATCTTGGAGGAGCG 1285  
DB 1055 GCAAGTGTAGCTTGGCGCGCTCCAGCGCGGAGCTTCAAGATCGCGGCCCGCGCG 996  
QY 1286 TGCTGCGGAGCGGAGCA 1303  
DB 995 TGGAGTGAACGCAAGA 978

Search completed: March 6, 2004, 02:20:54  
Job time: 158 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2004, 12:06:17 ; Search time 6189 seconds  
(without alignments)

18803.186 Million cell updates/sec

Title: US-10-021-657-7

Perfect score: 3897

Sequence: 1 gaattccaagcagggccctt.....gcgagtaaacgacgaattc 3897

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	976.8	25.1	1002	29	CG344416
2	939.2	24.1	991	29	CG047035
3	890.8	22.9	1070	29	CC889743
c 4	860	22.1	882	29	CC635363

c 5	858.6	22.0	887	29	CG067339
c 6	858	22.0	880	29	CG047034
c 7	857	22.0	879	29	CG685050
c 8	855.6	22.0	864	29	CG344408
9	844.6	21.7	915	29	CG535355
10	805.6	20.6	859	29	CG317335
c 11	802.2	20.7	887	29	CG611162
c 12	793.8	20.4	819	28	CC368220
c 13	779.2	20.0	848	29	CC723851
c 14	775.2	19.9	811	29	CG280923
c 15	774.2	19.9	997	11	AY110120
c 16	761.4	19.5	794	29	CG685043
c 17	751	19.3	773	28	CC368222
c 18	737.4	18.9	812	29	CG723843
c 19	721.6	18.5	771	29	CG610854
c 20	700.4	18.0	712	29	CG067338
21	694	17.8	733	29	CG371423
22	693	17.8	905	29	CG086520
c 23	674.4	17.3	833	29	CG317324
c 24	657.4	16.9	681	29	CG371414
c 25	654.8	16.8	748	28	BZ532472
c 26	651.6	16.7	700	29	CG086519
c 27	603.8	15.5	615	29	CG611171
c 28	512.4	13.1	517	10	AW424821
c 29	506	13.0	676	28	BZ798912
30	461.2	11.8	534	29	CG273559
31	460.4	11.8	563	10	AW519943
32	454	11.6	502	29	CG252095
33	450.2	11.6	474	28	BH872488
34	448	11.5	499	28	BZ798910
c 35	389.4	10.0	412	28	BH872489
36	386	9.9	585	12	BI075273
37	385	9.9	509	12	BI099541
38	384.2	9.9	606	10	BE494080
c 39	383.6	9.8	399	29	CG368180
40	375.8	9.6	672	29	CG257531
c 41	375.4	9.6	866	29	CG252417
42	372	9.5	649	12	BG948366
43	357.6	9.2	518	12	BG948686
44	356.4	9.1	578	12	BI075441
45	344.8	8.8	720	13	BQ840694

## ALIGNMENTS

RESULT 1  
CG344416/c  
LOCUS  
DEFINITION  
CG344416  
1002 bp DNA linear GSS 26-AUG-2003  
genomic survey sequence.  
ACCESSION  
CG344416  
VERSION  
GSS  
KEYWORDS  
GSS:34261682  
SOURCE  
Zea mays  
ORGANISM  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 1002)  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Numburg,A., Robbins,D. and Lakey,N.  
TITLE  
Consortium for Maize Genomics  
JOURNAL  
Unpublished (2002)  
COMMENT  
Other GSSs: OGI018TH  
Contact: Cathy Whitelaw  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.  
Location/Qualifiers  
1. .1002  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="ZM0.7.1.5\_KB"  
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

ORIGIN

Query Match 25.1%; Score 976.8; DB 29; Length 1002;  
Best Local Similarity 99.1%; Pred. No. 9.1e-151;  
Matches 993; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

QY 1082 GAGCACCATGGAGAACTCAGATCAGCGCGGACGCCATCGCCATCTTCCCACTAGC 1141  
Db |||||  
QY 1002 GAGCACCATGGAGAACTCAGATCAGCGCGGACGCCATCGCCATCTTCCCACTAGC 943  
Db |||||  
QY 1142 AGGSCCTCACAGTACATCGGCTCCCTCGTGTCTCTCATGGATCCTGGTCCAGAG 1201  
Db |||||  
QY 942 AGGSCCTCACAGTACATCGGCTCCCTCGTGTCTCTCATGGATCCTGGTCCAGAG 893  
Db |||||  
QY 1202 GTGAGCCTGAGGAGCAGAAAGCCCGAGATCATGGCCAGTCACTGGTCAACGGTGA 1261  
Db |||||  
QY 882 GTGAGCCTGAGGAGCAGAAAGCCCGAGATCATGGCCAGTCACTGGCAGACGGTGA 823  
Db |||||  
QY 1262 GCAGCTGAGAACTACACCGGATGACGACTGGCTTGGGTACTCTGTCAGGACAG 1321  
Db |||||  
QY 822 GCAGCTGAGAACTACACCGGATGACGACTGGCTTGGGTACTCTGTCAGGACAG 763  
Db |||||  
QY 1322 GACAGTGACCGTGCAGATGCGGTTCATCTCTACACCTACATCGCTGACCGGTGAATG 1381  
Db |||||  
QY 762 GACAGTGACCGTGCAGATGCGGTTCATCTCTACACCTACATCGCTGACCGGTGAATG 703  
Db |||||  
QY 1382 CGAGCATGCTCAAGATCAACTTCACTACAAATACCCCAAGTAAATGACCTGAACCTCACT 1441  
Db |||||  
QY 702 CGAGCATGCTCAAGATCAACTTCACTACAAATACCCCAAGTAAATGACCTGAACCTCACT 643  
Db |||||  
QY 1442 GATGTTTCAGTCTTCGAAATCAGAGCTGAAAGTGAATGCAATGCTGCTGAAACACCGTGT 1501  
Db |||||  
QY 642 GATGTTTCAGTCTTCGAAATCAGAGCTGAAAGTGAATGCAATGCTGCTGAAACACCGTGT 593  
Db |||||  
QY 1502 AGGGAATCGTGTACAGATCCTACATGACGCTGCTCTCGGTGACGGCATCTTCAACGGCG 1561  
Db |||||  
QY 582 AGGGAATCGTGTACAGATCCTACATGACGCTGCTCTCGGTGACGGCATCTTCAACGGCG 523  
Db |||||  
QY 1562 ACGGAGCTGTGGAGAACGACGAGGAGAGCGGAGTTCGAGTTCGCTCCAGAAC 1621  
Db |||||  
QY 522 ACGGAGCTGTGGAGAACGACGAGGAGAGCGGAGTTCGAGTTCGCTCCAGAAC 463  
Db |||||  
QY 1622 TGAGGATTTTCAGCGCATTTGTTTCAGAGAGTACTCCCTGAAAGCTGCGGGTACTGA 1681  
Db |||||  
QY 462 TGAGGATTTTCAGCGCATTTGTTTCAGAGAGTACTCCCTGAAAGCTGCGGGTACTGA 403  
Db |||||  
QY 1682 GCAGGATCCAAAGGACGAAAGTGTGGACATGACAGGTGAG--ATCACTGCTCCCTTG 1739  
Db |||||  
QY 402 GCAGGATCCAAAGGACGAAAGTGTGGACATGACAGGTGAGATCACTGCTGCTCCCTTG 343  
Db |||||  
QY 1740 CCAATTGCCACATGAGATTTCAACTGACAGACGAGAGTACTCTTGCCTGAGTCAAGAAC 1799  
Db |||||  
QY 342 CCAATTGCCACATGAGATTTCAACTGACAGACGAGAGTACTCTTGCCTGAGTCAAGAAC 283  
Db |||||  
QY 1800 TTTACATGAGATGACGCTGACTCCATCTGCAAGTGTGGTTCGCGGTTCGAGATCGGCA 1859  
Db |||||  
QY 282 TTTACATGAGATGACGCTGACTCCATCTGCAAGTGTGGTTCGCGGTTCGAGATCGGCA 223  
Db |||||  
QY 1860 CGCTGTGCGCGGATCTCCCGGAGAACAGCTTCGCGGAGCGGCTTCGATGCCCGCAACATCA 1919  
Db |||||  
QY 222 CGCTGTGCGCGGATCTCCCGGAGAACAGCTTCGCGGAGCGGCTTCGATGCCCGCAACATCA 163  
Db |||||

QY 1920 TCGTACGCTGCGGTTCATCGACCCCTGTGGCGCATCAAGAGTTCCTTCCACGTGGGT 1979  
Db |||||  
QY 162 TCATCACGCTGCGGTTCATCGACCCCTGTGGCGCATCAAGAGTTCCTTCCACGTGGGT 103  
Db |||||  
QY 1980 CAGAGGCCCTCTTACGCGCAGAGCATCAAGTCTGTGACGAGTTCACCTACAGGTGATCC 2039  
Db |||||  
QY 102 CAGAGGCCCTCTTACGCGCAGAGCATCAAGTCTGTGACGAGTTCACCTACAGGTGATCC 43  
Db |||||  
QY 2040 GCCGAGGAGGAGCGGAGATCGTGCAGGCCCGGCGCAGCGCA 2081  
Db |||||  
QY 42 GCCGAGGAGGAGCGGAGATCGTGCAGGCCCGGCGCAGCGCA 1

RESULT 2  
CG047035 991 bp DNA linear GSS 19-AUG-2003  
PUFWN667D ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBta0749KL2,  
genomic survey sequence.  
CG047035  
CG047035.1 GI:33919215  
GSS.  
Zeas mays  
Zeas mays  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 991)  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and  
Bennetzen, J.  
Maize Genomics Consortium  
Unpublished (2003)  
Other GSSs: PUFWN667B  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TP  
Class: sheared ends.

FEATURES  
source  
Location/Qualifiers  
1. .991  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBta0749KL2"  
/clone\_lib="ZM\_0.6\_1.0\_KB"  
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
Cot selected genomic DNA library"

ORIGIN

Query Match 24.1%; Score 939.2; DB 29; Length 991;  
Best Local Similarity 98.4%; Pred. No. 1.4e-144;  
Matches 970; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

QY 842 TGGTTGGCAGATCACAAAAGGAAACACAAAGGCTAAGCTCTTCTTGTTCGGAGTTAG 901  
Db 1 TGGTTGGCAGATCACAAAAGGAAACACAAAGGCTAAGCTCTTCTTGTTCGGAGTTAG 60  
QY 902 GTCAGGACACCATATGAATGAAGAAATCTTAA--TTTGGGGTCAACCAAGATTGTCTC 960  
Db 61 GTCAGGACACCATATGAATGAAGAAATCTTAAATTTGGGGTCAACCAAGATTGTCTC 120  
QY 961 TCTCGAGGTTGGGGGTCCCTAAGTTGGTAGTACCAATACCAATATATACCTAACAA 1020  
Db 121 TCTCGAGGTTGGGGGTCCCTAAGTTGGTAGTACCAATACCAATATATACCTAACAA 180  
QY 1021 ACCCAATCCATGCTACATACATACATAGCATCCATCTTGTAGACTGGACCTTCATCA 1080  
Db 181 ACCCAATCCATGCTACATACATACATAGCATCCATCTTGTAGACTGGACCTTCATCA 240



```

QY 930 TCTTAATTTGGGTCACACCAAGATTGTCTCTCTCGAGTTGGGGGGTCCCTAAGTTGG 989
Db 733 TCTTAATTTGGGTCACACCAAGATTGTCTCTCTCGAGTTGGGGGGTCCCTAAGTTGG 792
QY 990 TAGTAGCAATACCC- AATATATACCTTAACAAACCCATCCATCTACATACATACATAG 1048
Db 793 TAGTAGCAATACCC- AATATATACCTTAACAAACCCATCCATCTACATACATACATAG 852
QY 1049 CATCATCATCTGTAGACTGACCCCTTCCATCAAGAGCACCATGGAGGAAGCTCACAAC 1108
Db 853 CATCATCATCTGTAGACTGACCCCTTCCATCAAGAGCACCATGGAGGAAGCTCACCITCAC 912
QY 1109 GCCGGCAGCGCATCGCCATCTTCCCACTAGCAGGGCTCACAAGATACATCGGCTCCT 1168
Db 913 GCCGGCAGCGCATCGCCATCTTCCCACTAGCA- GGCTTCCCAAGTACATCGGCTCCT 971
QY 1169 CCTGTTCTCTCTCATGGATCCCTGTCAGAGTGGAGCTGAGGAGCAGGAAGGCC 1228
Db 972 TCTGTTGTCCTCTCATGGATCCCTGTCAGAGTGGAGCTGAGGAGCAGGAAGGCC 1031
QY 1229 GAGATCATGGCCAGTCAATCGGTGCAACGGTGGAGCAGCTG 1268
Db 1032 GAGATCATGGCCAGTCAATCGGCGC- ACGGTGGAGCAGCTG 1070

RESULT 4
CC635363/3
LOCUS OGKAW88TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBma0334P07,
DEFINITION genomic survey sequence.
ACCESSION CC635363
VERSION CC635363.1 GI:32013253
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 882)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
TITLE
JOURNAL
COMMENT Other_GSSs: OGKAW88TH
Contact: Cathy Whitelaw
7912 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
source Location/Qualifiers
1..882
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBma0334P07"
/clone_lib="ZM 0.7 1.5 KB"
/notes="vector: pBSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 22.1%; Score 860; DB 29; Length 882;
Best Local Similarity 99.2%; Pred. No. 1.6e-131;
Matches 875; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 903 TCAGGGACACATATGAATGAAGAAATCTTAATTTGGGTCACACCAAGATTGTCTCTC 962
Db 882 TCAGGGACACATATGAATGAAGAAATCTTAATTTGGGTCACACCAAGATTGTCTCTC 823

```

```

QY 963 TCAGGTTGGGGTCCCTAAGTTGGTAGCAATACCAATATATCACCTAACAAAC 1022
Db 822 TCAGGTTGGGGTCCCTAAGTTGGTAGCAATACCAATATATCACCTAACAAAC 763
QY 1023 CCAATCCATCTCATACATACATACATACATACATACATACATACATACATACAT 1082
Db 762 CCAATCCATCTCATACATACATACATACATACATACATACATACATACATACAT 703
QY 1083 AGCACCATGAGGAAGCTCACAATACGCGGCGAGCCATCGCATCTCTTCCCACTAGCA 1142
Db 702 AGCACCATGAGGAAGCTCACAATACGCGGCGAGCCATCGCATCTCTTCCCACTAGCA 643
QY 1143 GGGCTTCCCAAGTACATCGGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1202
Db 642 GGGCTTCCCAAGTACATCGGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 583
QY 1203 TGGAGCTTGGAGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1262
Db 582 TGGAGCTTGGAGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 523
QY 1263 CAGCTGAGGAATACCAACCGGATGCGAGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 1322
Db 522 CAGCTGAGGAATACCAACCGGATGCGAGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 463
QY 1323 ACAGTACCGTGCAGATCGCGTTCACTTCCATACACCTTACATCGCTGACCGGTTGAATGTC 1382
Db 462 ACAGTACCGTGCAGATCGCGTTCACTTCCATACACCTTACATCGCTGACCGGTTGAATGTC 403
QY 1383 GAGCATGTCTCAAGACTAATCTTCAATATCCGAGTAAATGACCTGAACTCACTG 1442
Db 402 GAGCATGTCTCAAGACTAATCTTCAATATCCGAGTAAATGACCTGAACTCACTG 343
QY 1443 ATGTTGAGTCTTCGGAATTCAGAGCTGAAAGCTGAAATGCTGCTGCTGCTGCTGCTGCTGCT 1502
Db 342 ATGTTGAGTCTTCGGAATTCAGAGCTGAAAGCTGAAATGCTGCTGCTGCTGCTGCTGCTGCT 283
QY 1503 GGAATCGTGTACAGATCTTACATGAGCTGCTCTCGGTGACGGCATCTTCAACGCGGA 1562
Db 282 GGAATCGTGTACAGATCTTACATGAGCTGCTCTCGGTGACGGCATCTTCAACGCGGA 223
QY 1563 CGCGAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1622
Db 222 CGCGAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 163
QY 1623 GAGGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1682
Db 162 GAGGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 103
QY 1683 CCAGGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
Db 102 CCAGGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 43
QY 1741 CATTGCCAATGAGCATTTTCACTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1782
Db 42 CATTGCCAATGAGCATTTTCACTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1

RESULT 5
CC635363/3
LOCUS FUICF65TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBta0550K10,
DEFINITION genomic survey sequence.
ACCESSION CC635363
VERSION CC635363.1 GI:33939519
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 887)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,

```











```
/db_xref="taxon:4577"
/clone="ZM068D11"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      21.7%; Score 844.6; DB 29; Length 915;
Best Local Similarity 98.0%; Pred. No. 5.6e-129;
Matches 897; Conservative 0; Mismatches 14; Indels 4; Gaps 4;

QY 308 GATTAGTGAATTAATTCAGCAATATCTTGTGTTGACATTTATATATATGTTGGGTGAG 367
Db 1 GATTAGTGAATTAATTCAGCAATATCTTGTGTTGACATTTATATATATGTTGGGTGAG 60
QY 368 ACGGATTAATATCATCATCAGAGAGCTTTATCTTCACTGCTCTCTGATTTGGTTTCAGA 427
Db 61 ACGGATTAATATCATCATCAGAGAGCTTTATCTTCACTGCTCTCTGATTTGGTTTCAGA 120
QY 428 TCATCTCTTCAAGTTCACAAAGATTTCTCAGTTTGGTTCAGTAAATTTTGAAGTGAG 487
Db 121 TCATCTCTTCAAGTTCACAAAGATTTCTCAGTTTGGTTCAGTAAATTTTGAAGTGAG 180
QY 488 GTTCTCTTAAATTCATATGCTTCTCTTCTTCTAGCTAGCAACTGCATGACTTTTCA 547
Db 181 GTTCTCTTAAATTCATATGCTTCTCTTCTTCTAGCTAGCAACTGCATGACTTTTCA 240
QY 548 CTTTGGGTTTCAAAATGACTTCACAGAAACAAATTCACCTTTTGGGTTTCAAAATTCCT 607
Db 241 CTTTGGGTTTCAAAATGACTTCACAGAAACAAATTCACCTTTTGGGTTTCAAAATTCCT 300
QY 608 CTTTCAAGATGATCTTTTCACTTGAATCTGATATAGGAACAAAGGAATGGCTCAGTTT 667
Db 301 CTTTCAAGATGATCTTTTCACTTGAATCTGATATAGGAACAAAGGAATGGCTCAGTTT 359
QY 668 TAAGGAACAATGATCAGATTTTCACTTTCAGAACTCTTCTGTTGGTTGAGTTTCAGACTT 727
Db 360 TAAGGAACAATGATCAGATTTTCACTTTCAGAACTCTTCTGTTGGTTGAGTTTCAGACTT 419
QY 728 TTTGTACCAAGCTGATGATCACAATATCTTGTTCCTTCCAAAGTCTGATTAACAGAACTGGCA 787
Db 420 TTTGTACCAAGCTGATGATCACAATATCTTGTTCCTTCCAAAGTCTGATTAACAGAACTGGCA 479
QY 788 ACTCTTAATTTGATTAATA-AAAGAAATAAATACAGTATCAGATATCATNTTCTTGGTT 846
Db 480 ACTCTTAATTTGATTAATAAAGAAATAAATAACAGTATCAGATATCATNTTCTTGGTT 539
QY 847 GGCAGATCACAAAAGGAACAAAGGCTAAGCCCTCCTACTTGTTCGGGAGTTAGGTGAG 906
Db 540 GGCAGATCACAAAAGGAACAAAGGCTAAGCCCTCCTACTTGTTCGGGAGTTAGGTGAG 599
QY 907 GGACACCATATGAATGAAGAAATCTTTAAATTTGGGGTCAACCAAGATTTGTCTCTCGA 965
Db 600 GGACACCATATGAATGAAGAAATCTTTAAATTTGGGGTCAACCAAGATTTGTCTCTCGA 659
QY 967 GGTTCGGGGTCCCTTAGTGTGGTAGCAATACCCCAATATATATCATCTTAACAAACCCAA 1026
Db 660 GGTTCGGGGTCCCTTAGTGTGGTAGCAATACCCCAATATATATCATCTTAACAAACCCAA 719
QY 1027 TCCATGCTACATACATACATGATCCATCATCTTGTAGCTTGGACCTTTCATCAAGAGCA 1086
Db 720 TCCATGCTACATACATACATGATCCATCATCTTGTAGCTTGGACCTTTCATCAAGAGCA 779
QY 1087 CCATGGAGGAAGCTCAATACACGCGGCGAGCCCATGCCCATCTTCCCACTAGCAGGGC 1146
Db 780 CCATGGAGGAAGCTCAATACACGCGGCGAGCCCATGCCCATCTTCCCACTAGCAGGGC 839
QY 1147 CTCACAGTACATCGCGCTCC-TCCTGGTGTCTCTCATGGAT-CTTGGTTCAGAGGTG 1204
Db 840 CTCACAGTACATCGCGCTCCCTTCTGTGTTGTCTCTCATGGATCCCTGGTCCAGAGGTG 899
QY 1205 GAGCCTGAGGAAGCA 1219
|||||
```

```
Db 900 GAGCCTGAGGAAGCA 914

RESULT 10
CG317335
LOCUS
DEFINITION
CG317335
genomic survey sequence.
ACCESSION
CG317335
VERSION
CG317335.1 GI:34234601
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 859)
REFERENCE
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSs: OG0CS18TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
FEATURES
Location/Qualifiers
1..859
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM068D11"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
Query Match      20.7%; Score 805.6; DB 29; Length 859;
Best Local Similarity 98.3%; Pred. No. 1.5e-122;
Matches 814; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1805 ATGAGGATGACGCTGACTCCATCTCGAGGTTGGGTTTCGGGTCGAGATCGGACGCTG 1864
Db 4 ATGAGGATGACGCTGACTCCATCTCGAGGTTGGGTTTCGGGTCGAGATCGGACGCTG 63
QY 1865 TCSCCGGATCTCCCGAGAACAGCTTCGGGCGAGCGCTTCGATCCCGCCAAACATCATCGT 1924
Db 64 TCSCCGAGATCTCCCGAGAACAGCTTCGGGCGAGCGCTTCGATCCCGCCAAACATCATC 123
QY 1925 ACCTCGGTTTCATCCACCGCTGTGGCGATCAAGAGTTTTCACGTCGGGTGAG 1984
Db 124 ACCTCGGTTTCATCCACCGCTGTGGCGATCAAGAGTTTTCACGTCGGGTGAG 183
QY 1985 GCCTCTCTAGCGCAGAGCATCAAGCTCGTGAGAGCTTCACCTACAGCGTGTCCGCCG 2044
Db 184 GCCTCTCTAGCGCAGAGCATCAAGCTCGTGAGAGCTTCACCTACAGCGTGTCCGCCG 243
QY 2045 AGGAAGCCGAGATCTCGAGGCCCGGCGAGCAAGAGAGGTACGTGCACAT 2104
Db 244 AGGAAGCCGAGATCTCGAGGCCCGGCGAGCAAGAGAGGTACGTGCACAT 303
QY 2105 GACTGTTTCGATTTCTTCAGTTTCATCGTCTTGGCCGGATGGACCTGATCTGATTGATTA 2164
Db 304 GACTGTTTCGATTTCTTCAGTTTCATCGTCTTGGCCGGATGGACCTGATCTGATTGATTA 363
QY 2165 TATATCCGTGTGACTTGTGAGGACAAATTAATGGGAGATGAACAGACATCTCTGTC 2224
|||||
```

Db	364	TATATCATGTGACTTGTGAGGGCAAAATTAAATATGGCGAGATGAAGCACGACATCCTGTG	423
QY	2225	ACGTTTCATCGAGCTAGCGAGGCCGCGACGAGCGGCGCGCTTCGGGACACGACGAG	2284
Db	424	ACGGTTTCATCGAGCTGGCGGAGGCCGCGACGAGCGGCGCGCTTCGGGACGATAGAG	483
QY	2285	CCTCCGGGACGTGTGTCTCAACTTCGTGTATCGCGCGCGGACACGACGCGGACGCGT	2344
Db	484	CCTCCGGGACGTGTGTCTCAACTTCGTGTATCGCGCGCGGACACGACGCGGACGCGT	543
QY	2345	GTGTGTGTTCAACGACATGGCCATGTCCCAACCCGAGCGTGGCGGAGAACTGCGCGCGGA	2404
Db	544	GTGTGTGTTCAACGACATGGCCATGTCCCAACCCGAGCGTGGCGGAGAACTGCGCGCGGA	603
QY	2405	GCTGTGGCGGTTTGAGCGGAGCGCGCGCGAGGAGGCGTGCCTGTGTCCTGCGG	2464
Db	604	GCTGTGGCGGTTTGAGCGGAGCGCGCGCGAGGAGGCGTGCCTGTGTCCTGCGG	663
QY	2465	CGGCGGTGACGCCGACGAAAGGGTTTCGCCGCCCGCGTGGCGAGTTGCGGGGCTTCCT	2524
Db	664	CGGCGGTGACGCCGACGAAAGGGTTTCGCCGCCCGCGTGGCGAGTTGCGGGGCTTCCT	723
QY	2525	CACCTACGACAGCCTCGGCGAAGCTGTCTACCTCCAGCCCTGCGTCAACGAGACGTCGG	2584
Db	724	CACCTACGACAGCCTCGGCGAAGCTGTCTACCTCCAGCCCTGCGTCAACGAGACGTCGG	783
QY	2585	CCGTGATCCCGCGCTCCCTCAGGTGAGCGCGCCGACACCGGACCTCC	2632
Db	784	CCGTGATCCCGCGCTCCCTCAGGTGAGCGCGCCGTCAGCTACGACC	831

RESULT 11				
CC611162/c				
LOCUS	CC611162	887 bp	DNA	linear
DEFINITION	OCLB788TH_ZM_0.7_1.5_KB_Zea mays genomic clone ZMMBma0315P08,			
	genomic survey sequence.			

KEYWORDS	CC000000.1	GU151972583
SOURCE	GSS.	
ORGANISM	Zea mays	
	Zea mays	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.	
REFERENCE	1 (bases 1 to 887)	
AUTHORS	Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.	
TITLE	Consortium for Maize Genomics	
JOURNAL	Unpublished (2002)	
COMMENT	Other_GSSS: OGLB98TV	
	Contact: Cathy Whitelaw	

```

Class: snaired ends.
FEATURES             Location/Qualifiers
     source            1..887
     /organism="Zea mays"
     /mol_type="genomic DNA"
     /strain="B73"
     /db_xref="taxon:4577"
     /clone="ZMMBMa0315P08"
     /clone_lib="ZM_0.7_1.5_KB"
     /note="Vector: pBSCS; Site 1: HindII; 0.7-1.5 kb
           methylation filtered genomic DNA library"
ORIGIN
Query Match          20.6%; Score 802.2; DB 29; Length 887;
Best Local Similarity 96.8%; Pred. No. 5.4e-122;

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 819)  
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.

TITLE Maize Genomics Consortium  
JOURNAL Unpublished (2003)  
COMMENT Other\_GSSs: PUH096TD

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..819

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBTA496024"

/clone\_lib="ZM\_0.6\_1.0 KB"

/note="Vector: PCR4-ToPO, Site: 1: EcoRI; 0.6-1.0 kb high"

CoI selected genomic DNA library"

ORIGIN

Query Match 20.4%; Score 793.8; DB 28; Length 819;  
Best Local Similarity 98.9%; Pred. No. 1.3e-120;  
Matches 810; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

QY 975 GGTCCCTAAGTTGGTAGTACCAATACCAATATATACCTAACCAAAACCAATCATGCT 1034

DB 1 GATCCCTAAGTTGGTAGTACCAATACCAATATATACCTAACCAAAACCAATCATGCT 60

QY 1035 ACATACATACATAGCATCCATCACTGTAGATGAGCCCTTCATCAAGAGCACCATGGAG 1094

DB 61 ACATACATACATAGCATCCATCACTGTAGATGAGCCCTTCATCAAGAGCACCATGGAG 120

QY 1095 GAAGTCTACATCACCGCGGACGCGCATCGCCATTCCTCCACATAGCAGGCGCTCACAG 1154

DB 121 GAAGTCTACCTCACCGCGGACGCGCATCGCCATTCCTCCACATAGCAGGCGCTCACAG 180

QY 1155 TACATCGCGCTCCTCCTCGTGTCTCTCATGATCTCTGTCACAGGTGGAGCTGAGG 1214

DB 181 TACATCGCGCTCCTCCTCGTGTCTCTCATGATCTCTGTCACAGGTGGAGCTGAGG 240

QY 1215 AAGCAGAAAGCCCGAGATCATGCGCAGTCTCGGTGCAAGGTTGGAGCAGCTGAGAAC 1274

DB 241 AAGCAGAAAGCCCGAGATCATGCGCAGTCTCGGTGCAAGGTTGGAGCAGCTGAGAAC 300

QY 1275 TACCACCGGATGCACGATGCTGTGTGCGGTACCTGTCCAGGACAGGACAGTACCGCTC 1334

DB 301 TACCACCGGATGCACGATGCTGTGTGCGGTACCTGTCCAGGACAGGACAGTACCGCTC 360

QY 1335 GACATGCGGTTCATTCCTCACTACCTACATCGCTGACCGGTGAATGTCGAGCATGCTC 1394

DB 361 GACATGCGGTTCATTCCTCACTACCTACATCGCTGACCGGTGAATGTCGAGCATGCTC 420

QY 1395 AAGACTAACTTCACCAATTACCCCAAGTAAATGACCTGACCTGACCTGACCTGACCTG 1454

DB 421 AAGACTAACTTCACCAATTACCCCAAGTAAATGACCTGACCTGACCTGACCTGACCTG 480

QY 1455 CGGAAATCAGAGCTGAAGCTGAATCGAATGCTGCTGAACACCTGTAGGGAATCGTGTA 1514

DB 481 CGGAAATCAGAGCTGAAGCTGAATCGAATGCTGCTGAACACCTGTAGGGAATCGTGTA 540

QY 1515 CAGATCCTACATGACGCTGCTCCTCGGTGACGGCATCTTCAACCCGACGCGGAGCTGTG 1574

DB 541 CAGATCCTACATGACGCTGCTCCTCGGTGACGGCATCTTCAACCCGACGCGGAGCTGTG 600

QY 1575 GAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCCCTCCCAAGAACCTGAGGATTTTCAG 1634

DB 601 GAGGAAGCAGAGGAAGACGGCGAGTTTCGAGTTCCCTCCCAAGAACCTGAGGATTTTCAG 660

QY 1635 CCCCATTGTTTCAGAGAGTACTCCCTCAAGCTGCGGTATATCTAGCCAGGCAATCCAA 1694

DB 661 CCCCATTGTTTCAGAGAGTACTCCCTCAAGCTGCGGTATATCTAGCCAGGCAATCCAA 720

QY 1695 GGCAGGCAAAAGTTGGACATGCGAGTGCAG--ATCACTGCTCCCTTGCATTCGCAACAT 1752

DB 721 GGCAGGCAAAAGTTGGACATGCGAGTGCAGTATCACTGGTCCCTTGCATTCGCAACAT 780

QY 1753 GAGCATTTCAACCTGAGACACGAGAGCTACCTTGCCTGAT 1791

DB 781 GAGCATTTCAACCTGAGACACGAGAGCTACCTTGCCTGAT 819

RESULT 13

CC723851

OGWER88TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0561008,

genomic survey sequence.

ACCESSION CC723851

VERSION CC723851.1

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 848)

Whitehead, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,

Cieck, R.W., Nurnberg, A., Robbins, D. and Lakey, N.

Consortium for Maize Genomics

Unpublished (2002)

Other\_GSSs: OGWER88TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..848

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBMA0561008"

/clone\_lib="ZM\_0.7\_1.5\_KB"

/note="Vector: pBCSK-; Site: 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

Query Match 20.0%; Score 779.2; DB 29; Length 848;  
Best Local Similarity 96.5%; Pred. No. 3.3e-118;  
Matches 818; Conservative 0; Mismatches 28; Indels 2; Gaps 2;

QY 577 AACAAATTCACCTTTGGTTCACAAATTCCTCTTCAGGATGTACTTTTCACTTGAAGTGT 636

DB 848 AACAAATTCACCTTTGGTTCACACAAATTCCTCTTCAGGATGTACTTTTCACTTGAAGTGT 789

QY 637 CATGTATAGAACAGGAATGGCTCAGTTTTTAAGAAACAATGTA CAGATTTCACTTCAG 696

DB 788 CATGTATAGAACAGGAATGGCTCAGTTTTTAAGAAACAATGTA CAGATTTCACTTCAG 729

QY 697 AACTCTTTCT-GGTTGGTTGAGTTTCAGACTTTTTCAGACCTGATGATGATGATGATGAT 755

DB 728 AACTCTTTCTGGTTGGTTGAGTTTCAGACTTTTTCAGACCTGATGATGATGATGATGAT 669

QY 756 TTGTTTCCAAAGTCTGATACAGAACTGGCAACTCCTAATTGATAATA-AAAAGATAA 814  
 Db 668 TTGTTTCCAAAGTCTAATAACAGAACTGGCAACTCCTAATTGATAATAATAAGATAA 609  
 QY 815 AATACAGTATCAGATATCTCATTTCTTGGTGGGAGATCACAAGAAAGGAAACACAAAGGC 874  
 Db 608 AATACAGTATCAGATATCTCATTTCTTGGTGGGAGATCACAAGAAAGGAAACACAAAGGC 549  
 QY 875 TAAAGCCTCTCACTTCTGTTGGGAGTTAGTFCAGGACACCATATGATGAAGAAATCTTA 934  
 Db 548 TAAAGCCTCTCACTTCTGTTGGGAGTTAGTFCAGGACACCATATGATGAAGAAATCTTA 489  
 QY 935 ATTGGGGTCAACCAAGATTGTCTCTCTCGAGGTGGGGGTCCCTAAGGTGGTAGTA 994  
 Db 488 ATTGGGGTCAACCAAGATTGTCTCTCTCGAGGTGGGGGTCCCTAAGGTGGTAGTA 429  
 QY 995 GCAATACCAATATATACCTTAAACCAATCATCTACATACATACATACATACATACCA 1054  
 Db 428 GCAATACCAATATATACCTTAAACCAATCATCTACATACATACATACATACATACCA 369  
 QY 1055 TCACCTGTAGACTGACCCCTTCATCAAGAGCACCATGGAGAGCTCAGATCAGCCGGC 1114  
 Db 368 TCACCTGTAGACTGACCCCTTCATCAAGAGCACCATGGAGAGCTCAGATCAGCCGGC 309  
 QY 1115 GACCCATCGCATCTTCCACTAGCAGGCGCTCACAAGTACATCGGCTCCTCTGGT 1174  
 Db 308 GACCCATCGCATCTTCCACTAGCAGGCGCTCACAAGTACATCGGCTCCTCTGGT 249  
 QY 1175 TGTCTCTCATGATCTGTGTCAGAGTGGAGCTGAGCTGAGGAGCAGAAAGCCGAGATC 1234  
 Db 248 TGTCTCTCATGATCTGTGTCAGAGTGGAGCTGAGCTGAGGAGCAGAAAGCCGAGATC 189  
 QY 1235 ATGSCCACTCATCGGTGCAAGCGTGGAGCAGCTGAGGAGCTGAGGAGCAGAAAGCCGAGATC 1294  
 Db 188 ATGSCCACTCATCGGTGCAAGCGTGGAGCAGCTGAGGAGCTGAGGAGCAGAAAGCCGAGATC 129  
 QY 1295 GCTTGTGGGTACTGTACGGCAAGCAGGACAGTGCAGATCGCATCGCTTCACTTCTTA 1354  
 Db 128 GCTTGTGGGTACTGTACGGCAAGCAGGACAGTGCAGATCGCATCGCTTCACTTCTTA 69  
 QY 1355 CACCTACATCGTGCAGCGGTGATGTCGAGAGTGCCTCAAGACTAATTCACCAATTA 1414  
 Db 68 CACCTACATCGTGCAGCGGTGATGTCGAGAGTGCCTCAAGACTAATTCACCAATTA 9  
 QY 1415 CCCCAAGG 1422  
 Db 8 CCCCAAGG 1

RESULT 14  
 CG280923  
 LOCUS  
 DEFINITION 811 bp DNA linear GSS 25-AUG-2003  
 genomic survey sequence.  
 CG280923  
 VERSION  
 CG280923.1 GI:34195137  
 GSS.  
 SOURCE  
 Zea mays  
 ORGANISM  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 811)  
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlking,T.,  
 Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other GSSs: OGWL096TV  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA

FEATURES  
 source  
 location/Qualifiers  
 1..811  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0667024"  
 /clone\_lib="ZM 0.7 1.5 KB"  
 /note="Vector: pBCK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

## ORIGIN

Query Match 19.9%; Score 775.2; DB 29; Length 811;  
 Best Local Similarity 98.8%; Pred. No. 1.5e-117;  
 Matches 802; Conservative 0; Mismatches 8; Indels 2; Gaps 2;  
 QY 465 GTCCATCTAATTTTGAAGTGAGGTTCTTAAATTTCAATTTCATTATGCTTCTTTCTTAG 524  
 Db 1 GTCCATCTAATTTTGAAGTGAGGTTCTTAAATTTCAATTTCATTATGCTTCTTTCTTAG 60  
 QY 525 ACTAGCAACTGCATGACTTTTCACTTTGGGTTCACAAATTTGACTCACAAGAAACAAATT 584  
 Db 61 ACTAGCAATTTGCATGACTTTTCACTTTGGGTTCACAAATTTGACTCACAAGAAACAAATT 120  
 QY 585 CACTTTTGGGTTTCAAAATTTCTTTCAGGATGTACTTTTCACTTGAATGTCTATGTATATA 644  
 Db 121 CACTTTTGGGTTTCAAAATTTCTTTCAGGATGTACTTTTCACTTGAATGTCTATGTATATA 179  
 QY 645 GGAACAGGAATGCTCAGTTTTTAAGGAACAATGTACAGATTTCAATTTCAGAACTCTTT 704  
 Db 180 GGAACAGGAATGCTCAGTTTTTAAGGAACAATGTACAGATTTCAATTTCAGAACTCTTT 239  
 QY 705 CTGGTTGGTTGAGTTTTCAGACTTTTGTACCAAGCTGTAGTGCATCAATACTTGTTCCTA 764  
 Db 240 CTGGTTGGTTGAGTTTTCAGACTTTTGTACCAAGCTGTAGTGCATCAATACTTGTTCCTA 299  
 QY 765 AAGTCTGATAACAGAACTGGCAACTCCCTAATTGATAATA-AAAAGATAATAACAGTA 823  
 Db 300 AAGTCTAATAACAGAACTGGCAACTCCCTAATTGATAATAATAACAGTA 359  
 QY 824 TCAGATATCTCATTTTCTTGGTTGCAGATCACAAGAAAGAACACAAGGCTAAGCTCC 883  
 Db 360 TCAGATATCTCATTTTCTTGGTTGCAGATCACAAGAAAGAACACAAGGCTAAGCTCC 419  
 QY 884 TACTTGTTCGGGAGTTAGTCTCAGGACACCATATGAATGAAAGAAATCTTAATTTGGGT 943  
 Db 420 TACTTGTTCGGGAGTTAGTCTCAGGACACCATATGAATGAAAGAAATCTTAATTTGGGT 479  
 QY 944 CACACCAAGATTGTCTCTCGAGTTGGGGGTCCCTAAGGTTGGTAGTAGCAATACCC 1003  
 Db 480 CACACCAAGATTGTCTCTCGAGTTGGGGGTCCCTAAGGTTGGTAGTAGCAATACCC 539  
 QY 1004 AATATATCACCTTAACAAACCAATTCATGCTACATACATACATACATACATACCTTGA 1063  
 Db 540 AATATATCACCTTAACAAACCAATTCATGCTACATACATACATACATACATACCTTGA 599  
 QY 1064 GACTGGACCTTTCATCAAGAGCACCATGGAGAAAGCTCACAATCAGCGCGGAGCGCAATC 1123  
 Db 600 GACTGGACCTTTCATCAAGAGCACCATGGAGAAAGCTCACAATCAGCGCGGAGCGCAATC 659  
 QY 1124 GCCATTCTTCCACTAGCAGGCGCTCACAAGTACATCGCGCTCTCTCTGTGTGCTCTC 1183  
 Db 660 GCCATTCTTCCACTAGCAGGCGCTCACAAGTACATCGCGCTCTCTCTGTGTGCTCTC 719  
 QY 1184 ATGGATCTCTGTCAGAGTTGGAGCTTGAGGAAGCAGAAAGGCGCCGAGATCATGGCAGT 1243  
 Db 720 ATGGATCTCTGTCAGAGTTGGAGCTTGAGGAAGCAGAAAGGCGCCGAGATCATGGCAGT 779

